Reconstruction and Quality Control of a Genome-Scale Metabolic Model for Methylococcus capsulatus

Lieven, Christian

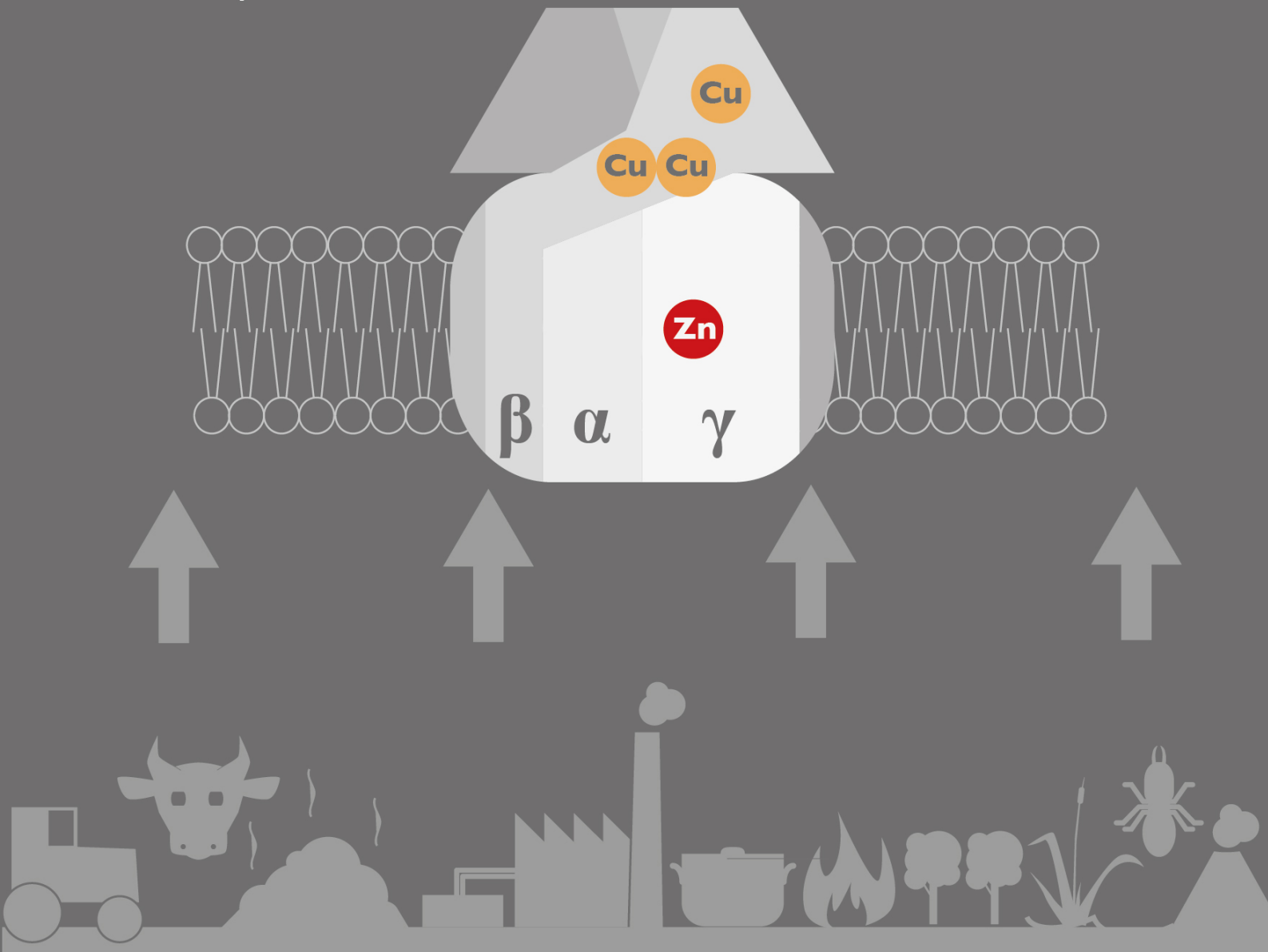
Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Reconstruction and Quality Control of a Genome-Scale Metabolic Model for *Methylococcus capsulatus*

Christian Lieven

PhD Thesis
February 2018



DTU Biosustain
The Novo Nordisk Foundation Center for Biosustainability

Reconstruction and Quality Control of a Genome-Scale Metabolic Model for Methylococcus capsulatus

PhD Thesis by Christian Lieven

February 2018

Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

Reconstruction and Quality Control of a Genome-Scale Metabolic Model for Methylococcus capsulatus

PhD thesis written by Christian Lieven

Title collage by Maja Rennig based on illustrations by Christian Lieven

Main Supervisor: Nikolaus Sonnenschein

Novo Nordisk Foundation Center for Biosustainability

Technical University of Denmark

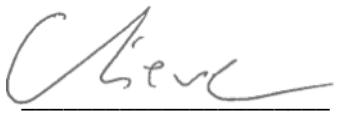
Kemitorvet, Building 220

2800 Kgs. Lyngby

Denmark

Preface

This thesis is written as a partial fulfillment of the requirements to obtain a PhD degree at the Technical University of Denmark. The work presented in this thesis was carried out between March 2015 and February 2018 at the Novo Nordisk Foundation Center for Biosustainability at the Technical University of Denmark, Kgs. Lyngby, and was supervised by Senior Scientist Nikolaus Sonnenschein. The Innovation Fund Denmark and the Novo Nordisk Foundation provided funding.



Christian Lieven

Kgs. Lyngby, February 2018

Abstract

Climate-change induced food scarcity is a grim prospect for the future of humanity. Alas, increasingly detrimental effects of global warming and overpopulation are predicted to beget this outlook by the year 2050.

Fortunately, the production of single-cell protein (SCP) from microbial fermentations presents an answer since it circumvents having to expand a spatially limited agriculture. It represents a controllable, scalable approach to food production that can be developed on a number of industrial waste products. The greenhouse gas methane is released abundantly into the atmosphere by both agricultural and industrial processes. To capitalize on this resource, the establishment of the protein-rich obligate methanotroph *Methylcococcus capsulatus* as an SCP product has been pursued. In addition, other methanotrophic organisms have been considered as cell factories for the production of chemicals from methane.

In this thesis, we recount the interplay between methane and the environment and outline the process of a genome-scale metabolic model (GEM) reconstruction. We then illuminate methods from software development that could streamline and quality control the reconstruction process, after which we continue to introduce existing genome-scale metabolic models of methanotrophs and their applications in biotechnology.

The main merit of this work lies in presenting memote (metabolic model tests), a set of community-curated test cases and the corresponding software, which enable the automated quality control of GEMs independent from reconstruction platforms or analyses software. Memote strongly supports publicly hosted and version controlled models, which we hope facilitates cross-community collaboration and research transparency.

Furthermore, we present the first manually curated GEM of *Methylococcus capsulatus*. The model comprises 750 genes, 877 metabolites and 898 reactions and

has been used to investigate the mode of electron transfer in this organism. Since it combines multi-layered biochemical and genomic information into one single knowledgebase, the availability of a GEM for *M. capsulatus* is an invaluable prerequisite for rational strain engineering towards an improved SCP production.

Dansk Resumé

Fødevareknaphed induceret af klimaændringer giver menneskehedens fremtid dårlige udsigter, og tiltagende skadelige følgevirkninger af global opvarmning og overbefolkning forudsiges at indtræde frem mod 2050.

Heldigvis udgør produktion af enkeltcelleprotein (SCP) fra mikrobielle fermenteringer en mulig løsning, da det omgår udfordringen med et rumligt begrænset landbrug. Det repræsenterer en kontrollerbar, skalerbar tilgang til fødevareproduktion, der kan baseres på en række industriaffaldsprodukter. Drivhusgassen metan frigives i atmosfæren ved både landbrug og industrielle processer. For at udnytte denne ressource er det forsøgt at etablere den proteinrige obligatoriske methanotroph Methylcococcus capsulatus som et SCP-produkt. Derudover er andre metanotrofe organismer blevet betragtet som cellefabrikker til fremstilling af kemikalier fra metan.

I denne afhandling gennemgås samspillet mellem metan og miljøet og konstruktionen af en genomskala metabolisk model (GEM) skitseres. Vi belyser derefter metoder fra softwareudvikling, der kan lette og kontrollere genopbygningsprocessen, hvorefter vi fortsætter med at introducere eksisterende metaboliske modeller for metanotrofer og deres anvendelser inden for bioteknologi.

Hovedværdien af dette arbejde ligger i at udviklingen af memote (metabolic model tests), et sæt af community-udviklede testcases og den tilsvarende software. Memote understøtter og opfordrer offentligt hostede og versionsstyrede modeller, som vi håber letter samarbejde mellem grupper og gennemsigtigheden af forskningen.

Desuden præsenterer vi den første manuelt kuraterede GEM af Methylococcus capsulatus. Modellen indeholder 750 gener, 877 metabolitter og 898 reaktioner og er blevet brugt til at undersøge elektrontransporten i denne organisme. Da en

GEM for M. capsulatus kombinerer biokemisk og genomisk information i en enkelt viden database, er tilgængeligheden af en sådan model en uvurderlig forudsætning for SCP-produktion.

Acknowledgements

In the short three years that I've spent as a PhD student, I've met a number of wonderful people here at the Center for Biosustainability, many of whom I've not only shared memorable experiences with but to whom I also owe a great deal of gratitude.

Before I begin, however, I ought to remark that none of this would have been possible if Birgitta Ebert had not made me aware of the job offer for this project in the first place. Thank you, Birgitta for being the initial impetus to these three truly exciting years!

Now, first of all I'd like to thank my main supervisor Nikolaus Sonnenschein not only for ultimately giving me the opportunity to pursue a PhD at the Center for Biosustainability, but also for enduring the many, probably very confusing, meetings on the complexities of metabolism in *M. capsulatus*. Your patience, trust, feedback and support has inspired me, and your approachable, jovial but sharp way motivated me to perpetually improve upon my own work. I thank you for giving me the opportunity to attend several international research conferences, exposing me to a great network of renowned experts.

Next, for granting me a chair at the table of the SIMS group, I'd like to thank Markus Herrgård. I greatly benefited from your drive and your vast experience and knowledge during these past three years. Being able to look to you for guidance has helped me immensely, particularly in my first year and on many occasions thereafter. I appreciate that you've taken the time to hear out my ideas, giving me new perspectives through feedback and critique.

I thank Leander Petersen, Sten Bay Jorgensen, Budi Juliman Hidayat, Subir Kumar Nandi, John Villadsen, Eleni Ntokou and Krist Gernaey for the countless hours of discussions about their experiences with *M. capsulatus*, their excitement about new results and their competent feedback.

My thanks go out to the DD-DeCaF team, Svetlana Kutuzova, Henning Redestig, Ali Kaafarni, Matyas Fodor and, most importantly, Mortiz E. Beber, for letting me observe and learn their software engineering skills, through following the same spring plans, collaborating on memote, and helping at the software carpentry course together.

For allowing me to visit his group on very short notice, I'd like to thank Jens Nielsen at Chalmers. I've had a great time and truly some of the most insightful discussions in my time as a PhD student. Parizad Babaei, Benjamín Sánchez Barja, Eduard Kerkhoven and Dimitra Lappa have been able to shed light on important user requirements for the development of memote, so by extension I'd like to express my gratitude to them too.

The warm welcome I received when I started at the SIMS group three years ago was tremendously relieving, for this, the excellent atmosphere ever since, and their support along the way, I'd like to thank Miguel Campodonico, Sumesh Sukumara, Emre Özdemir, Marta Matos, Ida Bonde, Kristian Jensen, Anne Sofie Lærke Hansen, Lars Schöning, Lasse Petersen and Joao Cardoso. In particular I'd like to thank Anna Koza, for the time we've spent during sampling runs and lab-equipment hauls, notwithstanding her patience and sense of professionalism when faced with difficulties.

To find an affordable place to live from abroad turned out to be more challenging than I anticipated. I am grateful for Solveig Siedler offering me to share her house for a month, and thus for being not only a great colleague but also an awesome flatmate. The same is true for Joao Cardoso, who I've shared with many musical and entertaining evenings.

Speaking of music and entertainment, I'd like to thank Kristian Jensen for forming a rock band with me two years ago and for sticking to it even though for the most part it was just he and I rocking it to a thin midi backing-track. I really appreciate your excitement and commitment to this little project of ours! Thank you for doing this with me from the start! At this point, I'd also like to thank the newcomers to this band: Christian Ravn, Christopher Knudsen, Nikolaus Sonnenschein and Carsten Beese. Our musical journey has been the ideal counterweight to my work as a PhD student.

To all my friends as well as present and past colleagues at the CfB, thank you for sharing all the ups and downs of moving from Hørsholm to DTU; water damages; summer, cocktail and Christmas parties, PhD Seminars and Friday Bars! It has been an absolute blast working and studying with you. Special thanks, in no specific order, go out to Anne Sofie Lærke Hansen, Ida Lauritsen, Mikkel Lindegaard, Matthew Jessop Fabre, Marta Matos, Steven Ocampo Duarte, Pasquale Colaianni, every single member of the SommerLab, Silvia Capucci, Kristian Jensen, George Borja, Joao Cardoso, Mafalda Cavaleiro, Patricia Maria Calero Valdayo, Alicia Jimenez Fernandez, Ivan Pogrebnyakov, Rosa Aragão Börner, Alicia Lis and essentially all the remaining Bactory people.

Finally I'd like to thank Maja Rennig. Every single thing in my life gets better when I can share it with you. Thank you for your unwavering support!

List of Publications

Included in this thesis:

1. Lieven, C., Herrgård, M. J., & Sonnenschein, N. (2018). Microbial C1 Metabolism: Recent Metabolic Modeling Efforts And Their Applications In Industrial Biotechnology. *Manuscript in Preparation.* **(Chapter 3)**

2. Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Babaei, P., Dräger, A., … Zhang, C. (2018). Memote: A community driven effort towards a standardized genome-scale metabolic model test suite. *Manuscript in Preparation.* **(Chapter 4)**

3. Lieven, C., Gernaey, K. V., Herrgard, M. J., & Sonnenschein, N. (2018). Modeling Methanotrophy: A genome-scale reconstruction of Methylococcus capsulatus. *Manuscript in Preparation.* **(Chapter 5)**

Not included in this thesis:

4. Cardoso, J. G. R., Jensen, K., Lieven, C., Hansen, A. S. L., Galkina, S., Beber, M., … Sonnenschein, N. (2017). Cameo : A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories. *bioRxiv*, *9*, 2013–2018. http://doi.org/10.1101/147199

Abbreviations

•OH	Hydroxyl Radical
AA	Amino Acids
ADH	Alanine Dehydrogenase
ANME	ANaerobic MEthanotrophs
API	Application Program Interface
ATP	Adenosine Triphoshpate
BCs	Biomass Compositions
BiGG	Biochemical Genetic and Genomic
BLAST	Basic Local Alignment Search Tool
BRENDA	Braunschweig Enzyme Database
CAD	Computer-aided Design
CBB	Calvin Benson Bassham cycle
CH_4	Methane
ChEBI	Chemical Entities of Biological Interest
CI	Continuous Integration
CO_2	Carbon Dioxide
COBRA	Constraint-based Reconstruction And Analysis
CoBSH	7-mercaptoheptanoylthreonine/ Coenzyme B
COMBINE	Computational Modeling in Biology Network
CoMSH	2-mercaptoethane sulphonate/ Coenzyme M
CSV	Comma-separated Values
CVCS	Centralized Version Control System
DHAP	Dihydroxyacetylphosphate
DNA	Deoxyribonucleic Acid
DVCS	Distributed Version Control System
DW	Dry Weight
EC	Enzyme Classification
EDD	Entner-Doudoroff Pathway
EGCs	Erroneous Energy-Generating Cycles
FA	Fatty Acids
FAD	Flavin Adenine Dinucleotide
FBA	Flux Balance Analysis
FBA	Fructose Bisphosphate Aldolase
FBC	Flux Balance Constraints
FLS	Formolase
FMN	Flavin Mononucleotide
GAM	Growth-Associated Maintenance
G_{ATM}	Atmospheric Growth Rate
GEM	Genome-scale Metabolic Model
GHG	Greenhouse Gas
GPR	Gene-Protein-Reaction
GRAS	Generally Recognized As Safe
GS/ GOGAT	Glutamine synthetase/ glutamine synthase

HTML	Hypertext Markup Language
ICM	Intracytoplasmic Membranes
ID	Identifier
InChI	International Chemical Identifier
IPA	International Phonetic Alphabet
KB	Knowledgebase
KDPG aldolase	2-Keto,3-Deoxy,6-Phosphogluconate Aldolase
KEGG	Kyoto Encyclopedia of Genese and Genomes
LPS	Lipopolysaccharide
MAT	MATLAB™ File Format
MCR	Methyl-Coenzyme M Reductase
MDH	Methanol Dehydrogenase
ME-Model	Model of Metabolism and Expression
MIRIAM	Minimal Information Required In the Annotation of Models
MMO	Methane Monooxygenase
MMOB	Methane Monooxygenase Regulatory Protein (Component of sMMO)
MMOH	Methane Monooxygenase Hydrolase (Component of sMMO)
MMOR	Methane Monooxygenase Reductase (Component of sMMO)
MNX	MetaNetX
N_2	Molecular Nitrogen
NAD/ NADH	Nicotinamide Adenine Dinucleotide
NC10	Unnamed candidate phylum pertaining to anaerobic oxidation of methane coupled to nitrite reduction
NGAM	Non Growth-Associated Maintenance
NH_4	Ammonium
NO_2	Nitrite
NO_3	Nitrate
NOAA	North Oceanic and Atmospheric Administration
O_2	Molecular Oxygen
ODS	OpenDocument Spreadsheet
ORF	Open Reading Frame
PDB	Protein Data Bank
PDF	Portable Document Format
PFK	Phosphofructokinase
PHA	Polyhydroxyalkanoates
PHB	Polyhydroxybutyrate
PMF	Proton Motive Force
pMMO	Particulate Methane Monooxygenase
QC	Quality Control
RAVEN Toolbox	Reconstruction, Analysis and Visualization of Metabolic Networks Toolbox
RBPC	Ribulose Bisphosphate Carboxylase
RCP	Representative Concentration Pathways
RNA	Ribonucleic Acid
RuBisCO	Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase
RuMP	Ribulose Monophosphate
SABIO-RK	System for the Analysis of Biochemical Pathways - Reaction Kinetics
SBML	Systems Biology Markup Language

SCP	Single-Cell Protein
SMILES	Simplified Molecular Input Line Entry System
sMMO	Soluble Methane Monooxygenase
SOP	Standard Operating Procedures
TA	Transaldolase
TCA	Tricarboxylic Acid Cycle/ Citric Acid Cycle/ Krebs cycle
TCDB	Transporter Classification Database
TDD	Test Driven Development
THMPT	Tetrahydromethanopterin
TSV	Tab-separated Values
UN	United Nations
US	United States
VCS	Version Control System
XLS/ XSLX	Microsoft Excel File Format
XML	Extensible Markup Language

List of Figures

List of Tables

Table of Contents

Introduction

Global warming and the accompanying climate change are unequivocal threats, the consequences of which humankind is already facing. Among many other effects, the oceans are acidifying and their levels rising, the occurrence of extreme weather such as storms, hurricanes, heat waves and draughts is increasing, and crop yields are shrinking (Field et al., 2014). As a consequence, simulations project that by 2050 climate-induced crop failure will likely contribute to food scarcity and subsequent malnourishment (Springmann et al., 2018).

In its 2017 revision of the world population prospects, the UN Department of Economic and Social Affairs predicts that by 2100 the global population will have grown to 11 billion from around 7.6 billion people (United Nations, Department of Economic and Social Affairs, 2017). Thus, in order to meet the demands of future generations the production of food and animal feed will have to be increased. However, this is associated with a number of challenges: 1) Traditional agriculture already occupies around 38% of the Earth's surface; further expansion will likely put natural habitats at risk. 2) Irrigation makes up 70% of the global freshwater use; further increase may compete with the availability of clean drinking water especially in dry regions of the planet. (Brauman et al., 2011) 3) According to the Fifth Assessment Report by the Intergovernmental Panel on Climate Change agriculture, forestry and other land use accounted for almost a quarter of the global greenhouse gas emissions in 2010 (Edenhofer et al., 2014); further intensification will likely increase this contribution. In short, food security and environmental sustainability seem to be opposing ideas when considering traditional agriculture, but what if we consider alternative sources?

Enter single-cell protein (SCP). The idea of using microbial biomass for human consumption as well as animal feed goes back to the start of the 20th century when it was suggested that residual brewer's yeast from beer production could be employed to boost protein content of animal feed. Two world wars and subsequent periods of supply shortages later, the process of growing *Saccharomyces*

cerevisiae aerobically on low-priced industrial by-products such as molasses or starch had been developed as a way of well-controlled, large scale feed and food production. This concept of waste-to-food remained successful until the 1980s, when political and agricultural changes allowed cheaper crops to dominate the global markets (Ugalde & Castrillo, 2002). Recently, the interest in this technology rekindled, however, now including algae and bacteria as possible sources of SCP in addition to fungi (Ritala, Häkkinen, Toivari, & Wiebe, 2017).

US natural gas prices have dropped significantly due to fracking and horizontal drilling and are predicted to remain low because of the expansive stores that are now accessible (Clemente, 2017). This creates the ideal conditions to economically prototype the use of specialized, gas-consuming microbes to produce both SCP and value-added compounds. One of these is *Methylococcus capsulatus*, which has attained generally recognized as safe (GRAS) status and grows aerobically on methane, the primary component of natural gas. Since methane is also an extremely potent greenhouse gas that is released abundantly from agricultural biogas facilities, landfills or petroleum flaring wells (Clomburg, Crumbley, & Gonzalez, 2017), this technology, unlike traditional agriculture, gracefully marries securing nutrition with climate sustainability.

In the first chapter of this thesis, I briefly illuminate the global impact of methane emissions and the flow of methane through the environment, before focusing on the remarkable enzymatic machinery that allow microbes such as *M. capsulatus* to use the gas as a source of carbon and energy. What follows is a glimpse into the metabolic diversity of aerobic methanotrophs, a thorough understanding of which is crucial for an efficient large-scale adoption of these organisms as biotechnological hosts.

Computer-aided design is applicable to biotechnology as much as it is to any other discipline of engineering. Hence, in chapter two, I refresh the key concepts of the digital toolset that is available for a hypothesis-driven creation of cell factories. I proceed to outline common approaches of reconstructing genome-scale metabolic models and close with introducing workflows from software development.

In chapter three, I review the existing literature with regards to the application of genome-scale metabolic models to either elucidate or enhance the native metabolism of methanotrophs.

The fourth chapter contains the manuscript of 'memote', a community-driven software solution that facilitates quality control and assessment of metabolic models. The tool heavily draws on the software development principles and reconstruction guidelines, delineated in the second chapter, to improve the way metabolic models are built and maintained.

In the final chapter, I present the manuscript of the genome-scale metabolic model of *Methylococcus capsulatus*. I hope that this contribution nudges forward the efforts of turning gas into optimal food or feed, and by this contributes, however much, to eventually reversing climate change and to ultimately prevent food scarcity and malnourishment.

Chapter 1. How To Breathe Methane And Live

1.1. The Gobal Role Of Methanotrophy

1.1.1. Methane as a potent green-house gas

Methane (CH_4) is the third most abundant greenhouse gas after water vapor and carbon dioxide (CO_2). Although the overall abundance of CH_4 in the atmosphere is lower than that of CO_2, projected over a period of 100 years the former is 28 times more powerful at trapping heat in the atmosphere, since its concentration decays at a much slower rate (Myhre et al., 2013). This results in a 20% contribution of methane to global warming since the year 1750 (Ciais, Sabine, & Bala, 2014). Its interaction with hydroxyl radicals (OH) and nitrogen oxide (NO_x) in the atmosphere further leads to self-aggravating effects, which prolong the atmospheric lifetime of the gas and consequently its impact.

The current, global average air mole fraction of CH_4 in the atmosphere (1803 ppb) has increased 2.5 fold as compared to the pre-industrial reference of 1750 (722 ppb). The primary reason for this rise in atmospheric methane is the increase of emissions originating from human activity, i.e. anthropogenic sources such as the combustion of fossil fuels. Depending on the prediction parameters (Figure 1-1A), it is projected that this value will double until the year 2100, unless appropriate actions are taken (Myhre et al., 2013). Global emissions of methane are estimated at 5.5×10^{11} kg per year (Kirschke et al., 2013). This is almost twice as much as the amount of biomass constituting the entire human population in 2005, which was calculated to sum to 2.87×10^{11} kg (Walpole et al., 2012).

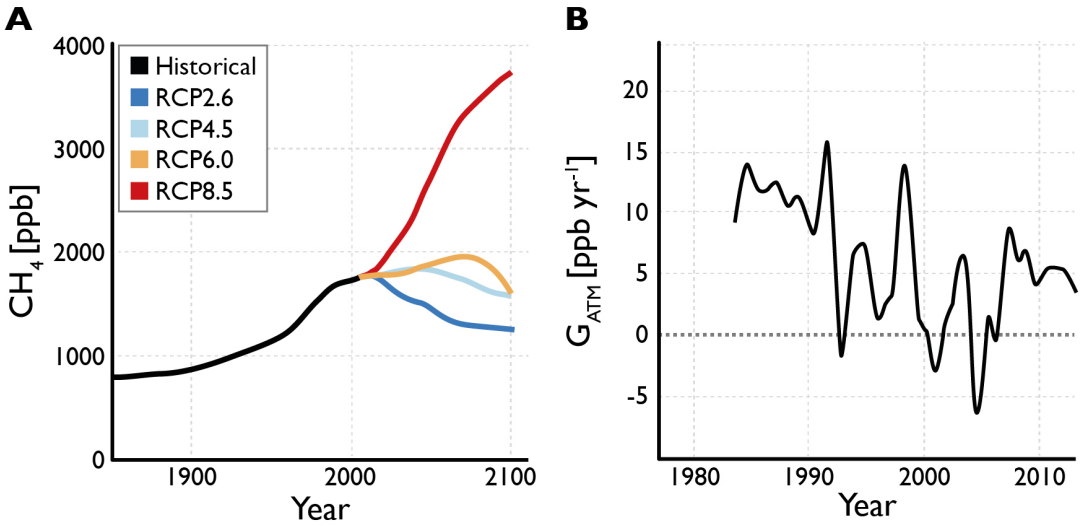


Figure 1-1. Trends of the global average methane levels in the atmosphere. A: Atmospheric methane concentration in parts per billion (ppb) in the period from 1850 to 2100. Measured values are displayed as the **black** line. Representative Concentration Pathways (RCP) are simulated scenarios based on socio-economic assumptions, which describe how future greenhouse gas (GHG) concentrations could affect the climate and what steps are necessary to maintain them at a certain level (Weyant, Azar, Kainuma, & Kejun, 2009). **RCP2.6** projects that a low limit of GHG concentrations is reached between the years 2010 and 2020, if GHG emissions are reduced by 70% and assuming all countries participate (van Vuuren et al., 2011). **RCP4.5** and **RCP6.0** assume peaks at 2040 and 2080, respectively, presupposing less stringent mitigation policies (Masui et al., 2011; Thomson et al., 2011). **RCP8.5** marks the worst-case scenario, a continuous increase until 2100 with little correction (Riahi et al., 2011). Figure adapted from Myhre et al. (2013). **B:** Atmospheric growth rate G_{ATM} in parts per billion (ppb) per year as measured by the North Oceanic and Atmospheric Administration (NOAA). Adapted from Saunois et al., (2016)

The difference between the amounts produced from CH_4 sources and the amounts consumed by sinks is the annual atmospheric growth rate. While the total amount of methane in the atmosphere has increased steadily since 1750, the growth rate has fluctuated drastically. Although exhibiting a general downward trend from 1985 until 2006, it has since started to increase again as shown in Figure 1-1B (Saunois et al., 2016). However, the global methane cycle is still not well understood and the estimates are associated with uncertainty (Kirschke et al., 2013). For instance, the impact of methane directly released by plants, animals or fungi is still controversial at best (Keppler, Hamilton, Braß, & Röckmann, 2006; Liu et al., 2015). Measurements and estimates are compounded by overlapping microbial activity, which has caused estimates to diverge by two orders of magnitude (Saunois et al., 2016). Beyond the uncertainties associated with

individual sources, there are records of systemic, abrupt warming events, which may drastically affect the balance of methane consumption and production (Hopcroft, 2017). To understand the robustness and the current budget of the global methane cycle, the sinks and sources and their individual impact are subject to ongoing research (Figure 1-2).

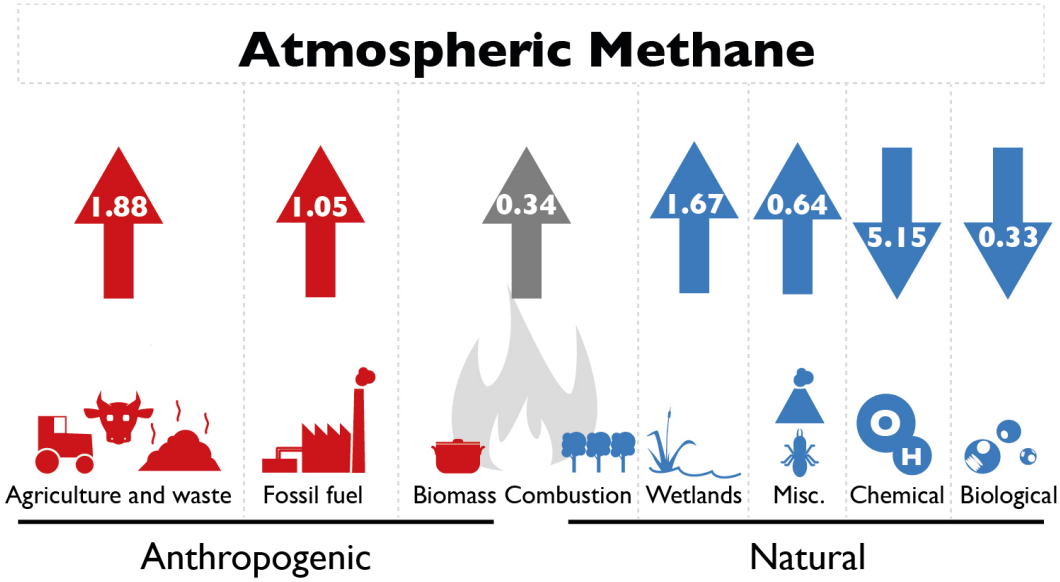


Figure 1-2. Sources and sinks of the global methane cycle. All values are in 10^{11} kg year^{-1} Natural sources, denoted by **blue** arrows pointing up, sum to 2.31 × 10^{11} kg year^{-1}. While sources related to human-activity, denoted by **red** arrows pointing up, account for 3.28 × 10^{11} kg year^{-1}. This includes the combustion of biomass (**gray** arrow) because the anthropogenic emissions far outweigh the natural emissions in top-down analysis by Saunois et al. (2016). Figure adapted from GlobalCarbonProject (2016).

1.1.2. Biogeochemical processes that release methane

As described by Saunois et al. (2016) there are three principal processes that create methane: thermogenic, biogenic and pyrogenic (Figure 1-3); the release of which can further be attributed to both natural and anthropogenic sources. 1) Fossilized deposits of organic material that have been buried deep inside the ground are broken down through the effects of pressure and compaction heat. The resulting thermogenic methane can reach the surface naturally by seeping through porous, marine or terrestrial rock, and volcanic action. Humans release it during the extraction, transportation, and refining of coal, natural gas, and oil. 2) Biogenic methane results from degradation of organic material carried out by *Archeabacteria*.

Not only do these organisms thrive in natural anaerobic environments such as waterlogged soils and sediments, but also the intestines of ruminants and termites. Landfills, water-treatment facilities, and some aspects of agriculture, specifically wetland rice fields and animal husbandry, comprise the manmade sources of biogenic methane. 3) Incomplete combustion of organic material emits pyrogenic methane. Here, natural sources are mostly wildfires, while anthropogenic sources include deliberate fire clearing of forests, the use of wood, peat, dung, agricultural residue and charcoal as fuel for open cook fires or domestic heating.

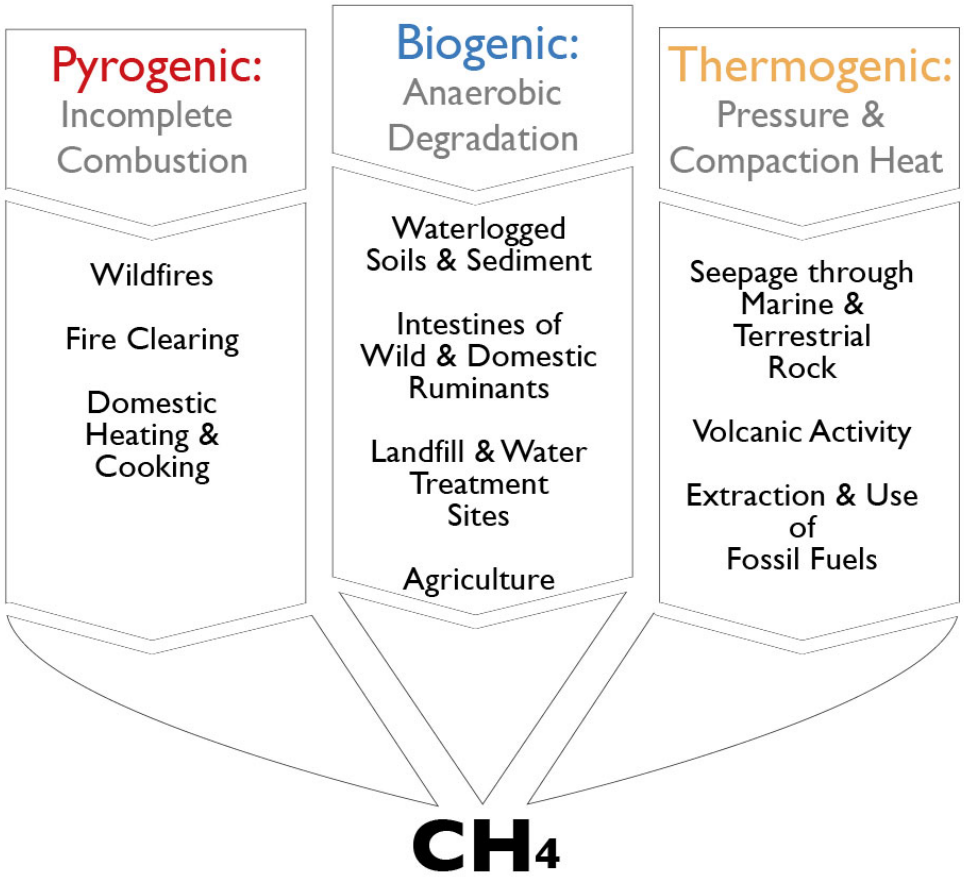


Figure 1-3. Overview of the three fundamental biogeochemical processes that produce methane. Associated sources, both natural and anthropogenic, are shown below each process respectively.

1.1.3. Global sources of methane

The total amount of methane emitted from anthropogenic sources at 3.28×10^{11} kg year^{-1} exceeds the natural emissions (2.31×10^{11} kg year^{-1}), which comprise the

emissions from wetlands (1.67×10^{11} kg year^{-1}) and from miscellaneous smaller sources such as seepage, volcanic activity, termites, lakes or oceans (0.64×10^{11} kg year^{-1}) (Saunois et al., 2016). Agricultural activity and waste management including sewage treatment, manure disposal, landfills, and rice cultivation make up the largest contributor to man-made CH_4 emissions with the production of 1.88×10^{11} kg year^{-1}. Increased growth of this source especially in tropical areas of the planet is assumed to have contributed the most to the increase of atmospheric methane since 2006. Processes of the fossil fuel industry are the second largest anthropogenic sources at 1.05×10^{11} kg per year. The remaining 0.34×10^{11} kg year^{-1} are released from the combustion of biomass as described above (Saunois et al., 2016).

1.1.4. Global sinks of methane

Opposed to the multitude of methane sources there are two principal types of sinks that together account for the removal of 5.15×10^{11} kg year^{-1} from the atmosphere: chemical and biological degradation (Saunois et al., 2016). The most significant chemical loss is through reaction with OH radicals in the tropo- and stratosphere:

(1) $$\boldsymbol{CH_4 + \cdot OH \rightarrow \cdot CH_3 + H_2O}$$

Subsequent reactions of the methyl radical with OH will first lead to the formation of formaldehyde, but ultimately to CO_2 and water vapour. Thus, the concentration of hydroxyl radicals in the atmosphere, i.e. the dominant buffer of rising methane emissions, is decreased by an increase in methane itself (Wuebbles & Hayhoe, 2002). This imbalance further leads to a feedback loop that ultimately contributes to the prolonged atmospheric lifetime of CH_4 mentioned at the beginning of this chapter (Myhre et al., 2013). Another pathway for chemical removal of CH_4 is through a number of reactions with radicals and chlorine gas, that are ultimately associated with the production of tropospheric ozone, stratospheric water vapour and CO_2, all of which further enhance the greenhouse effect (Wuebbles & Hayhoe, 2002).

The second sink is the biological oxidation of methane in dry soils capable of removing 0.33×10^{11} kg $year^{-1}$. Well-aerated, coarse-textured forest soils, which allow for a sufficiently high gas exchange by diffusion, exhibit maximal reaction rates. Methanotrophic bacteria, capable of using methane as their primary energy and carbon source, thrive in this environment and are responsible for driving this process (Smith et al., 2003). These organisms are adapted to the low atmospheric concentration of methane by expressing high-affinity enzymes that are able to break one of the strong C-H bonds of methane (Pratscher, Vollmers, Wiegand, Dumont, & Kaster, 2018).

To mitigate the impact of anthropogenic methane emission and to reduce the uncertainty in the interplay between sources and sinks, researchers began to investigate and exploit the unique metabolism of methanotrophs. Please refer to Chapter 3 for examples of recent applications involving methanotrophic microbes.

1.2. Breaking The Strongest Bonds

So far, methanotrophic organisms have been found to possess either one of two principal types of enzymes depending on whether they are adapted to the presence or absence of oxygen. The methyl-coenzyme M reductase (MCR) is expressed by anaerobic methanotrophs (ANME), while aerobic methane utilizers express the methane monooxygenase (MMO).

1.2.1. The methyl-coenzyme M reductase (MCR)

ANME are archaea that form syntrophic consortia with sulfate-reducing bacteria in anoxic deep-sea sediments (Conrad, 2009). The gas they consume is released from massive deposits of methane clathrates, which are mixtures of biogenic and thermogenic CH_4 molecules that form ice-like crystal structures with H_2O at specific temperatures and pressure. The amount of methane that is contained this way is estimated to range between 5000×10^{11} kg to 100000×10^{11} kg, the latter being around 3000 times as much as in the atmosphere currently. Despite the

considerable quantity of gas that is released from these stores globally (0.7×10^{11} kg year^{-1} - 3×10^{11} kg year^{-1}), the activities of ANME prevent it from reaching the atmosphere (Reeburgh, 2013). The MCR gene that is expressed by ANME is thought to catalyze the anaerobic oxidation of CH_4, while reducing a native heterodisulphide consisting of 7-mercaptoheptanoylthreonine also known as coenzyme B and 2-mercaptoethane sulphonate also known as coenzyme M (CoBS-SCoM) (Lawton & Rosenzweig, 2016a). Heterologous components of this process have first been identified in methane synthesizing archaea, which lead to the conclusion that the underlying mechanism behind anaerobic methanotrophy is essentially the reversal of methanogenesis (Timmers et al., 2017). The methyl-coenzyme M reductase of ANME differs from the MCR of methanogenic archaea in that it exhibits distinct post-translational modifications, requires a variant of the nickel tetrapyrrole cofactor F_{430}, and possesses a cysteine-rich area that is proposed to function as a redox-relay to activate the enzyme (Shima et al., 2012). See Figure 1-4 for a schematic view of the enzyme.

Since ANME in their natural habitat depend on synthrophic partner organisms to accept the electrons released by methane oxidation, a pure culture could not yet be established (Bennett, Steinberg, Chen, & Papoutsakis, 2018) and research on MCRs has been difficult (Lawton & Rosenzweig, 2016a). By cloning the gene from the metagenome of such a community, however, it was possible to express a functional ANME MCR in the methanogen *Methanosarcina acetivorans*, which marks a significant breakthrough in the study of this process (Soo et al., 2016). Heterologous expression of MCR in non-methanogenic hosts, however, is considered much more difficult as the genes responsible for the biosynthesis of the F_{430} cofactor need to be expressed in addition to identifying a suitable electron acceptor (Lawton & Rosenzweig, 2016a).

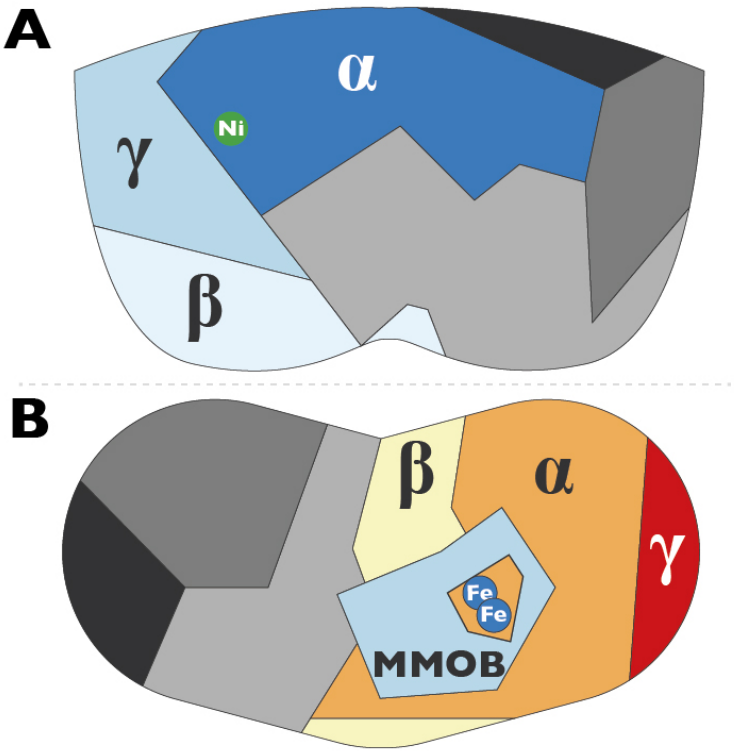


Figure 1-4 Schematic representation of MCR and sMMO (MMOH-MMOB). A: The methyl-coenzyme M reductase is a homodimer (**α, β, γ**)$_2$. The **α** subunit harbors the active site site, which contains the nickel tetrapyrrole F_{430} cofactor, here represented by a **green circle** labeled **Ni**. The corresponding PDB code of the structure is 1MRO. **B:** Like the MCR, the hydrolase component of the sMMO enzyme complex (MMOH) is also a homodimer (**α, β, γ**)$_2$. Here the complex of MMOH with the regulatory protein MMOB is shown. The **α** subunit of MMOH contains the active site, which is constituted by a diiron center. The individual iron atoms are depicted as **blue circles** labeled **Fe**. The corresponding PDB code of the structure is 4GAM. This figure has been simplified from Lawton & Rosenzweig (2016b).

1.2.2. The soluble methane monooxygenase (sMMO)

In contrast, methane monooxygenases, the key enzymes that facilitate aerobic methane oxidation, have been subjects of study since the 50s (C Anthony, 1983). MMOs exist in a soluble form (sMMO) and as a membrane-bound, so called particulate form (pMMO), which is most common among aerobe methanotrophs. In fact, there is only one genus of aerobe methanotrophs (*Methylocella*) that encodes only the sMMO, all other genera either contain only the pMMO or both (Camp et al., 2009). Broadly, either form of enzyme catalyzes the conversion of CH_4, oxygen and two reduction equivalents to methanol and water (Figure 1-5).

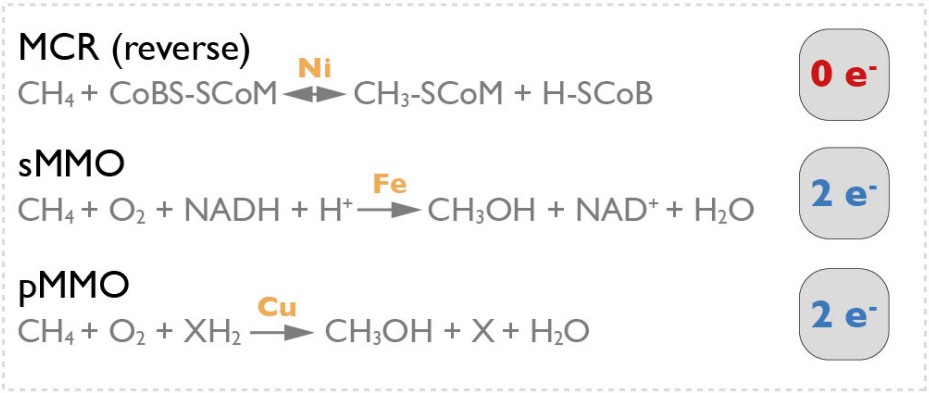


Figure 1-5. Reaction stoichiometry of MCR, sMMO and pMMO. All three reactions require different metal cofactors to activate the C-H bond in methane. The chemical **symbols** Ni, Fe, and Cu denote the metal cofactors nickel, iron, and copper, respectively. Anaerobic methane oxidation is catalyzed by the MCR. Hence, unlike the sMMO and the pMMO it does not require **2 electrons** to activate molecular oxygen. Figure adapted from Haynes & Gonzalez (2014).

A functional sMMO consists of three proteins MMOH, MMOR and MMOB, which interact with each other in a coordinated manner. The following mechanism as summarized by Lawton & Rosenzweig (2016) is the result of several studies and assumed to depict the process most accurately (Figure 1-6): A $(\alpha\beta\gamma)_2$ homodimer, MMOH, is the central entity of the reaction having an active site in each α subunit comprised of a diiron center. During the first step, the second entity, MMOR, is reduced with electrons transferred from NADH via FAD to internal Iron-Sulfur clusters, while the MMOH is in an oxidized state ($MMOH_{ox}$), its diiron(III) core still coordinating one oxygen atom from the previous iteration. The reductase MMOR then converts $MMOH_{ox}$ to $MMOH_{red}$, by binding to what is referred to as the 'canyon region' and reducing the diiron(III) site to diiron(II). This releases MMOR, now oxidized, from the `canyon region` and the oxygen atom dissociates in the form of H_2O. In the second step, the third entity (MMOB) now binds to the same `canyon region` effecting a conformational change in $MMOH_{red}$ (Figure 1-4). The diiron(II) core can now react with O_2, and, after achieving homolytic cleavage of the O-O bond via several intermediates, it arrives at a diiron(IV) state. Methane then binds to $MMOH_{red}$, and reacts with one of the oxygen atoms to form methanol during the

third and final step. This returns MMOH to the initial, oxidized state and releases MMOB from the 'canyon region', thus completing the cycle. The intricacy of this three-way mechanism involving MMOH, MMOR and MMOB likely complicates functional heterologous expression (Lawton & Rosenzweig, 2016a). So far, researchers succeeded in expressing sMMO from *Methylosinus trichosporium* OB3b in five Pseudomonas strains. While the recombinant host strains gained the ability to degrade trichloroethylene, the rate of the reaction was significantly slower than for the native *M. trichosporium* OB3b. The ability to grow on methane was not reported in this study (Jahng & Wood, 1994).

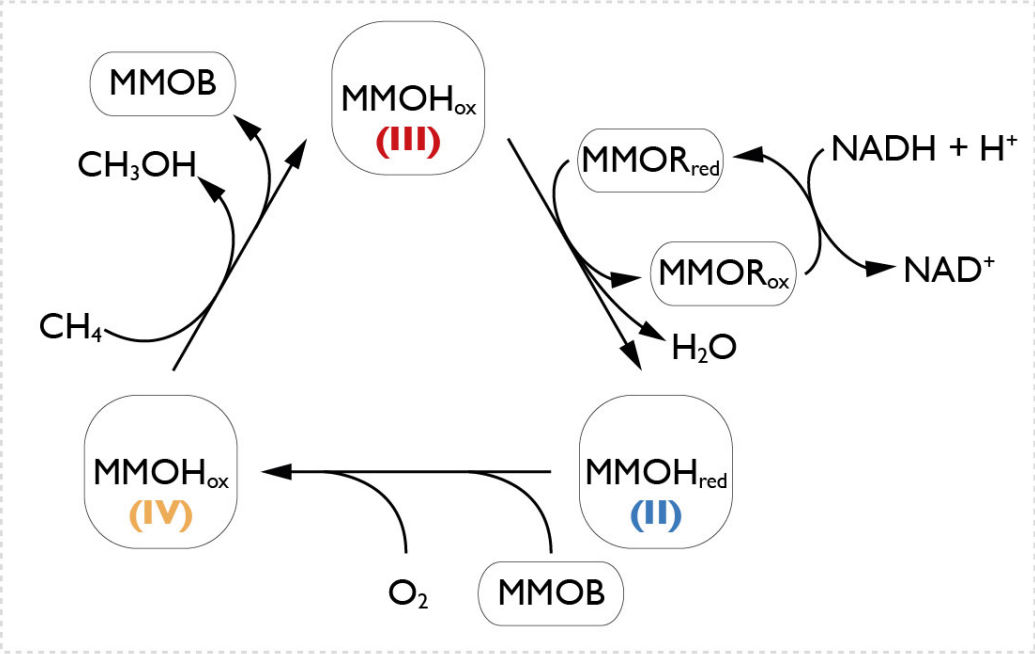


Figure 1-6. Catalytic reaction cycle of the sMMO. The main component, MMOH, contains a diiron cluster that iterates through three distinct oxidation states, denoted by **(III)**, **(II)**, and **(IV)**. This successively allows the binding of MMOR, then MMOB and molecular oxygen, and lastly methane. Simplified from Lawton & Rosenzweig (2016a).

1.2.3. The particulate methanemonooxygenase (pMMO)

Similar to how the MCR requires nickel and sMMO requires iron to be catalytically active, the pMMO relies on copper ions for its activity. The particulate methane monooxygenase is a homotrimer, each monomer consisting of the three subunits, α, β, and γ, encoded by the genes *pmoCAB*. Subunit β (*pmoB*) consists of

two cupredoxins, one at the N terminus and one at the C terminus, linked by two transmembrane helices. The helices anchor the subunit in the inner membrane such that the soluble cupredoxins are protruding into the periplasm. Evidence suggests that the N-terminal copper protein contains the active site, but the exact amount of copper that can be modeled to fit there differs between methanotrophs. For *Methylococcus capsulatus*, *Methylosinus trichosporium* OB3 and *Methylocystis sp.* M two copper ions can fit with the crystal structures of the corresponding enzymes, whereas for *Methylocystis sp.* Rockwell a model with just one copper ion agrees best (Lawton & Rosenzweig, 2016b). Specifically the β subunit of *M. capsulatus* has an additional, dedicated monocopper site, which is absent in the enzymes of the other mentioned organisms. Both, the α (*pmoA*) and γ (*pmoC*) subunits are multi-helix membrane proteins, which possess conserved sites that support metal-ion complexes. Using crystallography, copper or zinc was identified to bind to the γ subunit. Yet, since the pMMO could only be crystalized when excess copper or zinc was added, and since zinc is a known inhibitor of the pMMO, it is currently speculated that these sites don't actually bind metal ions, but instead their physiological role is to transfer protons. No metal ions have been observed interacting with the conserved site of subunit α (Lawton & Rosenzweig, 2016b). See Figure 1-7, for a schematic representation of the subunits and metal sites in the pMMO of *M. capsulatus*.

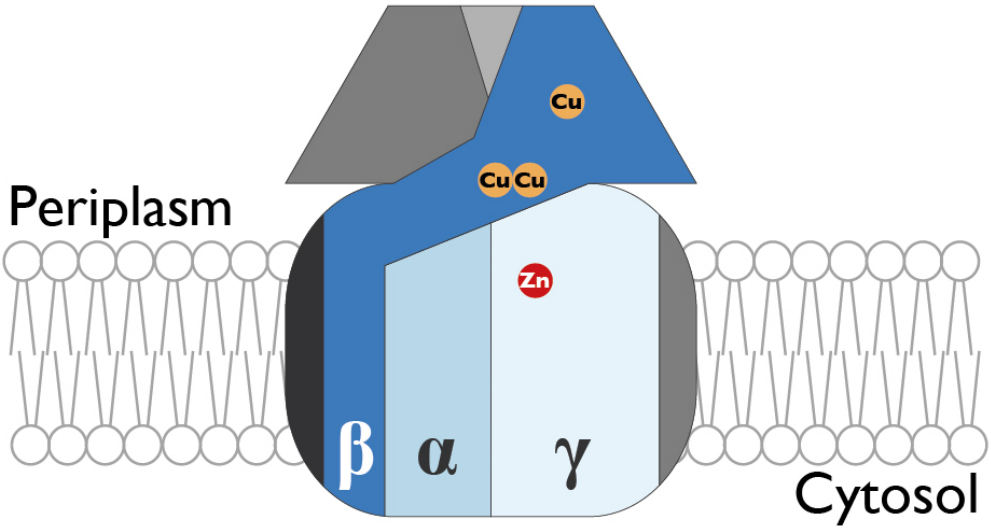


Figure 1-7. Schematic representation of the pMMO of *M. capsulatus*. The particular methane monooxygenase is a homotrimer (**α, β, γ**)$_3$. The **β** subunit consists of two cupredoxins that extend into the periplasm (diagonal section at the center and top) and two transmembrane helices (horizontal section, bottom left). The **β** subunit of the pMMO of *M. capsulatus* further contains a dicopper site, which constitutes the active site of the enzyme, and a separate monocopper site (**Orange circles** labeled **Cu**). The **α** and **γ** subunits are multi-helix membrane proteins. In crystallography experiments zinc has been found to occupy a conserved site on the **γ** subunits (**Red circle** labeled **Zn**). It is assumed that the physiological role of this site is associated with proton shuttling. The corresponding PDB code of the structure is 3RFR. Simplified from Lawton & Rosenzweig (2016b).

The study and heterologous expression of the pMMO is difficult because it is an integral membrane protein with a complex trimeric structure. In nature, it is expressed in an expansive stack of intracytoplasmic membranes (ICMs) with unknown function (Lawton & Rosenzweig, 2016a). Using the pMMO recombinantly is further compounded by its reliance on copper, which is toxic at high concentrations. On top of that, individual pMMOs differ with respect to the required amounts of copper, which has hampered the elucidation of the underlying enzymatic mechanism (Lawton & Rosenzweig, 2016b). While Gou et al. (2006) succeeded at expressing pMMO from *Methylosinus trichosporium* OB3b in recombinant *Rhodococcus erythropolis* LSSE8-1, the productivity was minimal.

Yet another complicated research question, the nature of electron donor to the pMMO, is discussed at length in Chapters 3 and 5 of this thesis.

Not only in structure and chemical mechanism but also in terms of enzyme performance are sMMO and pMMO different from each other. The sMMO has a broad substrate range that includes alkanes, alkenes, benzene, styrene, naphthalene, ethyl benzene and cyclohexane, as well as halocarbons such as the carcinogenic industrial solvent trichloroethylene (Sirajuddin & Rosenzweig, 2015), while the pMMO is much more limited: merely the preferential C-2 oxidation of a small number of straight chain alkanes and alkenes has been reported (Elliott et al., 1997). The substrate range of the MCR, however, seems to be limited exclusively to CH_3-SCoM and H-SCoB or CH^4 and CoBS-SCoM for the forward and reverse reaction, respectively, as other substrates have not been reported yet.

The fact that both sMMO and pMMO require two electrons (Figure 1-5) for the activation of oxygen, and the ensuing redox-neutral conversion of methane to formaldehyde reduce the reaction efficiency in comparison with the MCR. Since no energy is gained directly from the oxidation of CH_4, roughly one in three formaldehyde molecules are oxidized to CO_2, resulting in a 33% loss of carbon. In contrast to the subsequent carbon efficiency of 67%, anaerobic methanotrophs expressing a reversible MCR potentially offer a carbon efficiency of 100% (Lawton & Rosenzweig, 2016b).

Since isolated ANME have not yet been cultured, I will proceed to give an overview of the metabolism of aerobe methanotrophs exclusively.

1.3. Aerobe Methanotrophs And Their Key Differences

Gram-negative, aerobe bacteria that use methane as their primary source of carbon and energy were first described over a century ago and have since been studied extensively (Trotsenko & Murrell, 2008). As outlined in Chapter 1.1, these organisms play a large role in the reduction of atmospheric methane by being the

only biological sinks in the global methane cycle. Yet, their unique metabolism also holds potential for diverse biotechnological applications. Hence, the use of methanotrophs has recently been proposed as an economical means to reduce the immense anthropogenic greenhouse gas emissions (Clomburg et al., 2017).

Aerobe methanotrophs belong either to the phyla verrucomicrobia, NC10 or proteobacteria,. With some exceptions methanotrophic proteobacteria are mostly neutrophilic and mesophilic, while those that are affiliated with verrucomicrobia can grow at a pH lower than 1, and at temperatures close to 65ºC (Camp et al., 2009).

1.3.1. Verrucomicrobial and NC10 methanotrophs

Verrucomicrobial methanotrophs have only been isolated a little under a decade ago; consequently literature on their metabolism is still scarce. In addition to being morphologically different and possessing unique enzymes that utilize rare earth elements, they have been found to be autotrophs, assimilating carbon from CO_2 via the Calvin cycle while oxidizing CH_4 (Niftrik, 2014) or molecular hydrogen as a source of energy (Mohammadi, Pol, Alen, & Jetten, 2016). Like verrucomicrobia, methanotrophs belonging to the candidate phylum NC10 are also autotrophs, but are further able to use denitrification to anaerobically carry out methane oxidation i.e. use oxygen that is released internally to drive this reaction (Ettwig et al., 2010).

1.3.2. Alpha- and Gammaproteobacteria

Proteobacteria can further be sub-classified into gamma- and alpha-proteobacteria, also largely referred to as Type I, and II, respectively. In Type I methanotrophs, ICMs are arranged as collections of vesicular disks, as opposed to the paired peripheral layers of ICM that are present in Type II methanotrophs (Figure 1-8). This morphological difference further correlates with the predominant formaldehyde assimilatory pathway that is active in either group: the ribulose monophosphate (RuMP) pathway in Type I, and the serine cycle in Type II. Organisms belonging to the species of *Methylococcus* and *Methylocaldum* are

sometimes referred to as Type X methanotrophs, as they not only possess genes from the RuMP pathway and the serine cycle, but also a ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) from the Calvin cycle (Trotsenko & Murrell, 2008).

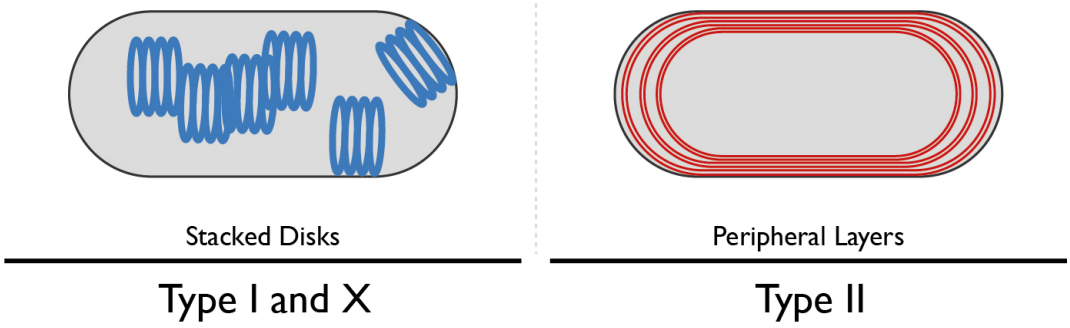


Figure 1-8 Schematic representation of the characteristic ICM arrangement of methanotrophs belonging to Alpha- or Gammaproteobacteria. Type I and X methanotrophs = gammaproteobacteria,; Type II = alphaproteobacteria. Redrawn from Dalton (2005).

Type II methanotrophs are of particular interest due to their tendency of storing excess carbon in fatty acids, polyhydroxyalkanoates (PHA), or polyhydroxybutyrate (PHB). Opposed to this, Type I and X methanotrophs are much more efficient in converting methane into biomass (Karthikeyan, Chidambarampadmavathy, Cirés, & Heimann, 2015). Because of this and the fact that they generally have high protein contents of up to 70% per weight (Pieja, Morse, & Cal, 2017), Type I methanotrophs have been favored for the production of single-cell protein (SCP).

In Chapter 3, I reference current applications of well-studied methanotrophs that have been translated into genome-scale metabolic models. In Chapter 5, I present a genome-scale metabolic model of the gammaproteobacterium, *M. capsulatus*, which has been a promising candidate for the efficient production of SCP on a large scale (Øverland, Tauson, Shearer, & Skrede, 2010).

Chapter 2. Fundamentals Of Genome-Scale Metabolic Reconstruction

2.1. Central Concepts Of Stoichiometric Modeling

2.1.1. Metabolic Network as S-Matrix

The use of computers to simulate complex systems of the modern world is ubiquitous: companies execute digital risk assessment to determine the stability of the financial system (Nyman et al., 2018), global agencies such as the UN rely on computational extrapolation of population data streams to advise on governmental policies (United Nations, Department of Economic and Social Affairs, 2017), meteorologists rely on climate models to accurately forecast a change of weather or the impact of climate change (Edenhofer et al., 2014; Hatfield, Subramanian, Palmer, & Düben, 2017), and engineers of all disciplines learn to thoroughly design and test constructions using computer-aided design (CAD) software before commencing work on any physical representations (Jensen, 2007).

Compared to the man-made systems mentioned above, a living cell is a system of equal if not greater complexity. Intracellular processes do not happen in isolation but are embedded into a network of components that operate simultaneously at different scales: genes synthesize proteins and proteins catalyze chemical reactions, which then influence both previous steps. Hence, in systems biology and biotechnology just like in other fields, researchers avail themselves of computational tools to study and analyze the behavior of cellular systems in the form of metabolic, signaling, or regulatory networks. Hereafter, I focus on metabolic networks, as they are immediately applicable to biotechnology and can be extended to include regulatory effects. Unless stated otherwise, the concepts presented herein have been adapted from Palsson (2015).

A prerequisite for the reconstruction of metabolic networks is an accurate account of metabolic components and their interactions. Since these interactions are of a chemical nature inherently, they can be expressed as chemical equations, which relate the conversion of compounds according to a fixed stoichiometry. From this follows that the mass and energy of the system is conserved, meaning that material cannot be generated from nothing and that the chemical equations are balanced.

For a given system, the relevant components of a metabolic network are metabolic enzymes, which can be represented by the chemical reactions they catalyze, and metabolites, which interact with other metabolites via those reactions. As an example, consider the three chemical equations below:

R1: $A \leftrightarrow B$

R2: $2B \longrightarrow C$

R3: $C \longrightarrow 2A$

Using the stoichiometric coefficient from each metabolite in each reaction, a stoichiometric matrix $\boldsymbol{S}$ can be constructed that accurately describes this set of reactions. Metabolites constitute the rows and reactions the columns of the matrix. Furthermore, metabolites that are consumed in a reaction have negative coefficients, while those that are produced have positive coefficients:

(2)
$$\boldsymbol{S} = \begin{array}{c} \\ \boldsymbol{A} \\ \boldsymbol{B} \\ \boldsymbol{C} \end{array} \begin{array}{c} \begin{array}{ccc} \boldsymbol{R1} & \boldsymbol{R2} & \boldsymbol{R3} \end{array} \\ \begin{pmatrix} -1 & 0 & 2 \\ 1 & -2 & 0 \\ 0 & 1 & -1 \end{pmatrix} \end{array}$$

The $\boldsymbol{S}$ matrix plays an important role in systems biology modeling, as it not only represents the fundamental topology of a metabolic network, but also allows the mathematical prediction of intracellular fluxes i.e. metabolic conversion rates with a few additional assumptions and constraints.

2.1.2. Steady-State and Various Constraints

The primary assumption is that the modeled system is in a steady state, meaning that there is no net accumulation of mass over time. The concentrations of metabolites can be considered time-invariant for instance when an organism grows exponentially in a continuous or batch culture. In that case, the accumulation of intracellular material is offset by the rate of duplication meaning there is no net change in intracellular metabolite pools when the point of reference is one single cell. This is also referred to as a pseudo steady state (or dynamic equilibrium).

With the vector $\boldsymbol{v}$ representing the fluxes through the reactions of the network that satisfy the steady state and physicochemical mass-balance criteria mentioned above, and interpreting the $\boldsymbol{S}$ matrix as a system of linear mathematical equations, the metabolic network can now be formulated as:

(3) $$\boldsymbol{Sv = 0}$$

The solution space of this system of linear equations is unbounded and contains an infinite number of possible solutions, so called steady-state flux distributions. To reduce the solution space and make it bounded, it is assumed that the flux through the network is subject to thermodynamic laws and finite enzyme capacity (Cotten & Reed, 2016). Hence, for all reactions that have been characterized to be irreversible in physiological conditions $\boldsymbol{I}$, the corresponding flux is constrained to be a non-negative number:

(4) $$\boldsymbol{v_j \geq 0 \; \forall \; j \in I}$$

Moreover, upper bounds $\boldsymbol{ub}$ and lower bounds $\boldsymbol{lb}$ imposed on all reactions $\boldsymbol{R}$ determine the capacity of a corresponding enzyme.

(5) $$\boldsymbol{lb \leq v_j \leq ub \;\; \forall \; j \in R}$$

For internal reactions, the bounds are chosen to be sufficiently large unless experimentally determined not to be. However, uptake or production rates can

generally be measured for controlled cultures in experiments . These rates can then be used to set the bounds of the corresponding reactions in the model, bringing the resulting flux distributions closer to an experimental reference state.

2.1.3. Flux Balance Analysis

In addition to the steady state assumption and additional flux capacity and reaction directionality constraints, it is often assumed that some form of selection pressure further governs the systemic behavior of cellular metabolism, essentially optimizing it towards a specific phenotype. One such optimization for instance, is rapid growth, which in a nutrient rich natural habitat allows prevailing against slow growing organisms. Another potential metabolic strategy is to reduce the overall energy required for growth at a certain rate, which grants an advantage over competitors when the carbon source is scarce requiring the cell to use it as efficiently as possible. The applicability of either growth objective is highly dependent on external factors, thus demanding careful consideration (Schuster, Pfeiffer, & Fell, 2008).

The most widely used constraint-based method is flux balance analysis (FBA), which implements the cellular objective as a mathematical optimization problem where the flux through one specific reaction v_{obj} is maximized or minimized:

$$\max\ v_{obj}\ s.t.\ (3), (4)\ and\ (5) \tag{6}$$

Since the flux through virtually any reaction in the metabolic network can be optimized this way, FBA is frequently used to determine the hypothetical maximal yields of compounds-of-interest. Moreover, by gradually iterating through a set of the constraints for individual reactions it is possible to study the network behavior in different conditions. For instance, one could gradually decrease the flux allowed through the oxygen uptake reaction while optimizing for fast growth, to study the metabolic shift from aerobic to anaerobic growth.

2.2. A genome-scale metabolic model

2.2.1. Manual Reconstruction

Building a metabolic network with a few reactions is fairly straightforward, however, constructing a metabolic network on a genome-scale generally requires a methodic approach that is both time- and labor-intensive. Broadly there are four sequential steps that lead towards a metabolic model of this scale, which are often reiterated to refine the reconstruction (Figure 2-1).

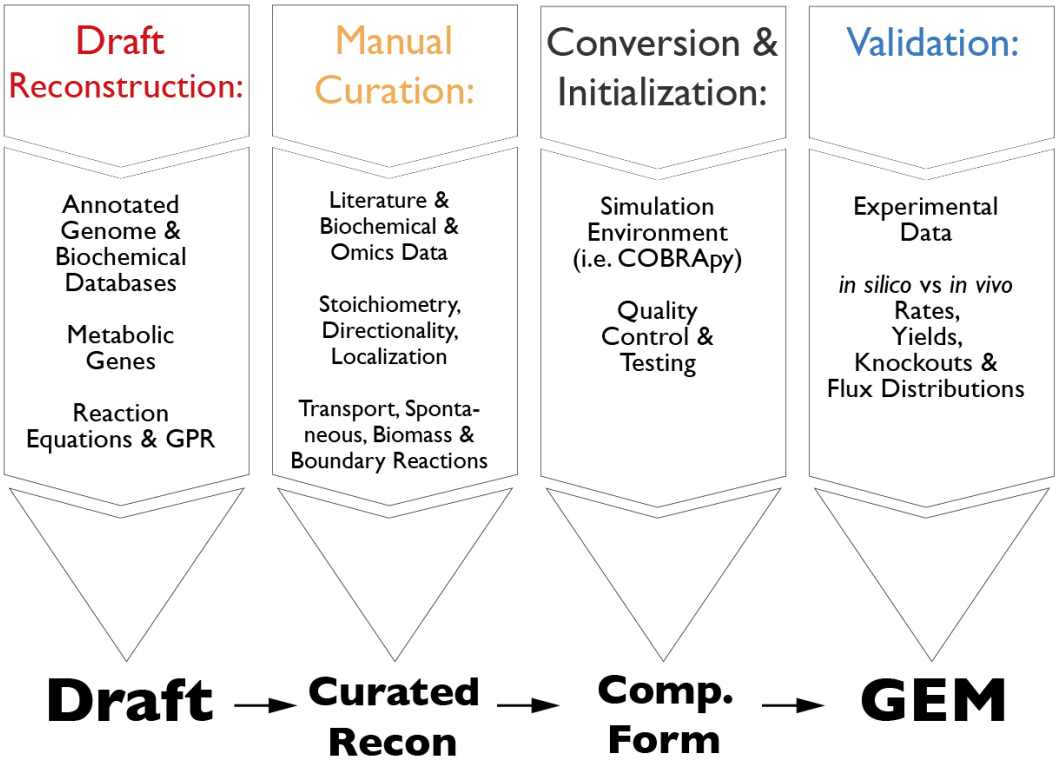


Figure 2-1. Overview of the manual reconstruction process for genome-scale metabolic models. Adapted from Palson (2015).

The first step is the reconstruction of a draft network. It is essential for this step to obtain a functionally annotated, full genome sequence of the organism of interest. From the sequence annotation it is then possible to enumerate all genes that encode for metabolic enzymes, or depending on the annotation quality, the chemical reactions associated with each gene-product, for instance through enzyme classification (EC) numbers. If EC numbers are not available, then the reactions can usually be looked up in comprehensive biochemical databases such

as KEGG (Kanehisa, Furumichi, Tanabe, Sato, & Morishima, 2017), UniProtKB (Bateman et al., 2017), BRENDA (Placzek et al., 2017 - www.brenda-enzymes.org) or Sabio-RK (Wittig et al., 2012).

It is important to also capture the association between genes and their products by composing gene-protein-reaction (GPR) rules, which help to determine whether two given genes are isozymes or part of a larger protein complex or whether the genes individually carry out a single or multiple functions (Figure 2-2). Hence, GPR rules provide a measure of confidence to a reconstruction by extending the list of reactions with an accurate representation of the genomic interconnectivity.

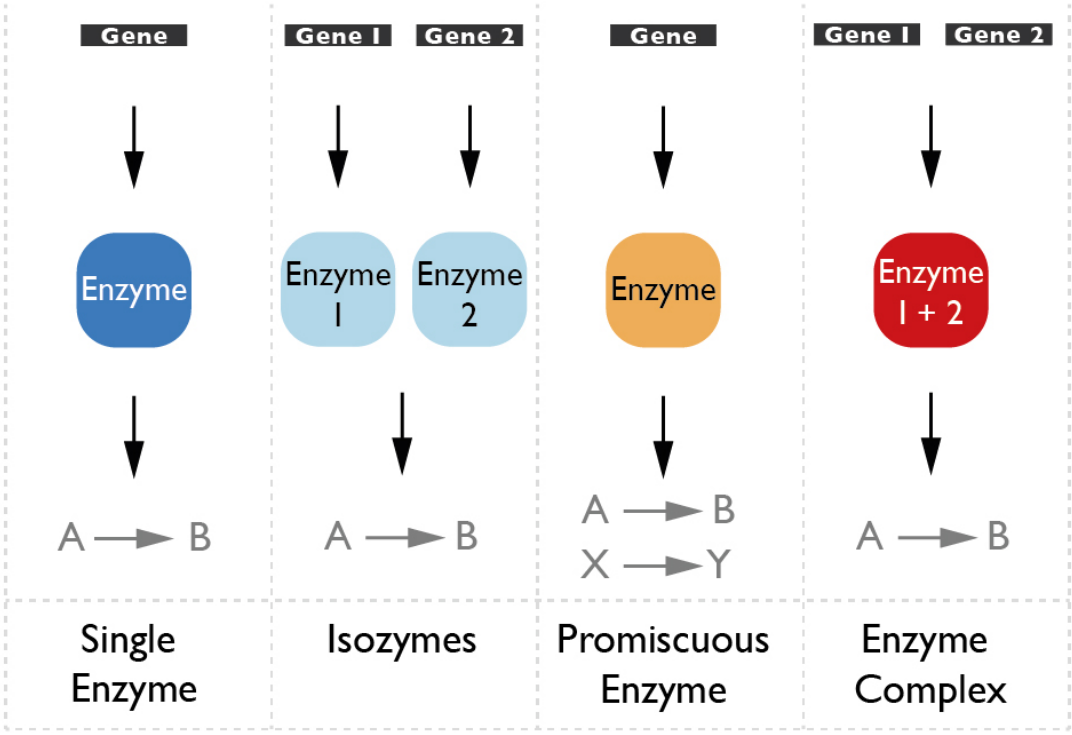


Figure 2-2. Gene-Protein-Reaction associations. Gene-Protein-Reaction rules are Boolean rules that translate these associations into machine-readable relations between genes and the corresponding reactions. For instance, a reaction ***A*** → ***B*** that is catalyzed by isozymes bears the GPR rule: Enzyme I *OR* Enzyme 2. If, however, an enzyme complex catalyzed the same reaction, the rule would be: Enzyme I *AND* Enzyme 2. Figure adapted from Machado, Herrgård, & Rocha, (2016).

Next comes the lengthy process of manual curation. Here, data from high-throughput characterization of primary cell components generated by functional genomics, proteomics, transcriptomics or other 'omics' methods, as well as

knowledge from scientific literature is combined to custom fit the often generic information from the genome annotation. Ideally, the network is manually curated by paying special attention to the types of co-factors used, charge- and mass-balancing, physiological directionality and intracellular location of each reaction. In this step, transport and spontaneous reactions are added and the system boundaries are defined. This requires the addition of specific, unbalanced pseudo-reactions that provide or remove metabolites from the metabolic network.

One such pseudo-reaction is the biomass reaction, which by definition simulates the drainage of metabolites and energy equivalents required to construct 1 gram dry weight of cells (Thiele & Palsson, 2010). The stoichiometry and components of this reaction are calculated from measurements of the overall biomass composition and the compositions of main cellular fractions (Figure 2-3). Experimental measurements may not allow a total coverage of the biomass composition; to a degree it is possible, however, to estimate certain parameters from the genome sequence, or infer them from a related organism. Yet, this likely affects the ability of the model to predict an optimal growth rate in FBA that corresponds to measured values of the organism of interest (Thiele & Palsson, 2010).

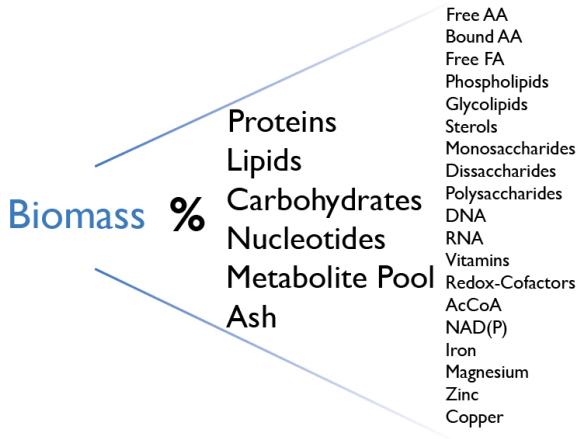


Figure 2-3. Categorical approach to the biomass composition. The measured biomass composition lays the foundation for the construction of a biomass reaction in

GEMs. The individual components of the biomass composition can be measured in a variety of ways from considering only macromolecular groups (e.g. proteins, lipids, mucleotides) to smaller groups (e.g. free AA, free FA, DNA) to individual metabolites (e.g. alanine, palmitic acid, adenosine). The biomass reaction of a GEM integrates the percent contributions of each component as stoichiometric coefficients normalized to the production of 1 g DW of Biomass (Thiele & Palsson, 2010).

The third step is the initialization of the reconstruction in a computable form, which includes importing the model in a simulation environment such as the COBRA toolbox (Heirendt et al., 2017) or COBRApy (Ebrahim, Lerman, Palsson, & Hyduke, 2013). To ensure that there have been no technical mistakes that could affect predictions with the model, preliminary quality control (QC) of the network with FBA is carried out at this stage. Since a genome-scale metabolic model can include thousands of components, it is likely that errors or inconsistencies are present. To ensure consistency, researchers for instance test if all biomass precursors can be synthesized, if the model adheres to the strict mass- and charge-balance requirements such that no metabolite can be produced from nothing, and if there are no loops that lead to any unnatural energy-generating cycles (Fritzemeier, Hartleb, Szappanos, Papp, & Lercher, 2017).

Lastly, the metabolic model is validated against organism-specific experimental data. Based on the available data, the network is constrained to the organism-specific reference condition (i.e. aerobic growth on methane) and an analogous experiment is carried out *in silico*. Depending on what type of data is available this ranges from a comparison of growth, energy or production yields, to evaluating differences between *in silico* and *in vivo* knockout studies. Thus, in addition to the previously mentioned quality control on the implementation of the network, validation constitutes a second level of QC, related to the accuracy of the model.

These four steps can be repeated as more information on an organism becomes available, which will increase the biochemical and genetic information reflected in the model. Essentially the model serves as a knowledgebase that integrates all available information on a specific organism and can be extended to include addition information, such as regulatory interactions, enzyme states or gene expression (O'Brien, Lerman, Chang, Hyduke, & Palsson, 2013).

2.2.2. Automated Reconstruction

Since the steps required to build a genome-scale metabolic model have been outlined precisely in the form of standard operating procedures (SOPs) (Thiele & Palsson, 2010), it has been possible to formalize them into software. The use of such automated reconstruction tools can help reduce the total time spent on the reconstruction process. Today, many automated reconstruction tools exist and some have been reviewed at great length elsewhere (Chang, 2018; Hamilton & Reed, 2014). Summarized briefly, the current tools vary quite drastically: Some are toolboxes consisting of individual scripts (Agren et al., 2013); others are fully integrated standalone applications (Karp et al., 2010) and still others are web-applications (Henry et al., 2010). They differ in the amount of steps in the reconstruction process they support and are often connected to specific biochemical databases such as KEGG (Kanehisa et al., 2014) or Metacyc (Caspi et al., 2014) to resolve genome annotations of the reconstructed organism into reaction information.

Although the dependency on specific databases may vary between reconstruction tools, building a draft by identifying reactions in a biochemical database is a common shared feature. While this removes the arduous task of iterating through all metabolic genes to identify associated reactions in databases manually, it may introduce systemic errors. The underlying databases typically contain generic reaction information based on data from well-studied model organisms. Hence for the more obscure, less-studied organisms, such as methanotrophs, the resulting draft networks require particularly thorough curation and gap filling, because enzyme function and cofactor usage may not be adequately represented in the databases. Furthermore, an automatically constructed biomass reaction is likely to misrepresent a specifically measured biomass composition of the organism of interest.

Since there is no adequate measure of quality or completeness for genome-scale metabolic models (J. Monk, Nogales, & Palsson, 2014), it is impossible to compare automated draft reconstruction tools based on their output. Instead,

their numerous individual features have to be weighed against one another. One also has to consider the fact that manually reconstructing a genome-scale metabolic model grants a more thorough understanding of the target organism than running an automatic pipeline.

In either case, however, the result is not only an expansive collection of organism-specific information, but also a predictive mathematical tool that can be interrogated to show systemic relationships. In Chapter 3 of this thesis, I review the recent literature on metabolic models of methanotrophy, and present such a genome-scale model for the obligate methanotroph *Methylococcus capsulatus* in Chapter 5.

2.3. Systems Biology 2.0

2.3.1. Version Control

Genome-scale metabolic models (GEMs) are comparable to software in a number of ways. Running FBA on a GEM is similar to executing a function in the way that in both settings there is a defined number of inputs and a certain expected output. The source code of a program can span thousands of lines that are functionally interconnected, which is equivalent to a large scale metabolic network where as many components are just as tightly interlinked. Moreover, both may be difficult to understand at a glance, which means that debugging is comparably hard. Lastly, building a GEM or writing a program happens in increments and both of them are often improved after release, thus several versions may be in circulation simultaneously.

In software engineering, special tools and methods have been developed that facilitate the management of software with regards to the above-mentioned similarities. One of them is the idea of a dedicated version control system, which tracks the changes made to specified files over time. Unless otherwise stated, the following information is adapted from Chacon & Straub (2014).

Having a detailed record of the incremental changes and a program to manage it, allows the developer to revisit earlier versions, by simply undoing the operations between any two given points in time. Statistics on the nature of changes can be kept and give insight into the timeline of the project. Specifically, the history of changes can help to pinpoint the introduction of bugs or to recover a previous instance after the loss of data.

Of course this set of changes can be kept locally, but when working collaboratively on larger software projects there are two common approaches: centralized (CVCS) and distributed version control systems (DVCS). In CVCS, like the name implies there is a central server that holds the files and history of the entire project. Contributors first update a local, working copy of the project with the most recent changes from the central server and then proceed to make recorded edits locally. When the work is done, they save their local changes by appending them to the central history of changes. An example for a CVCS is Subversion (https://subversion.apache.org/).

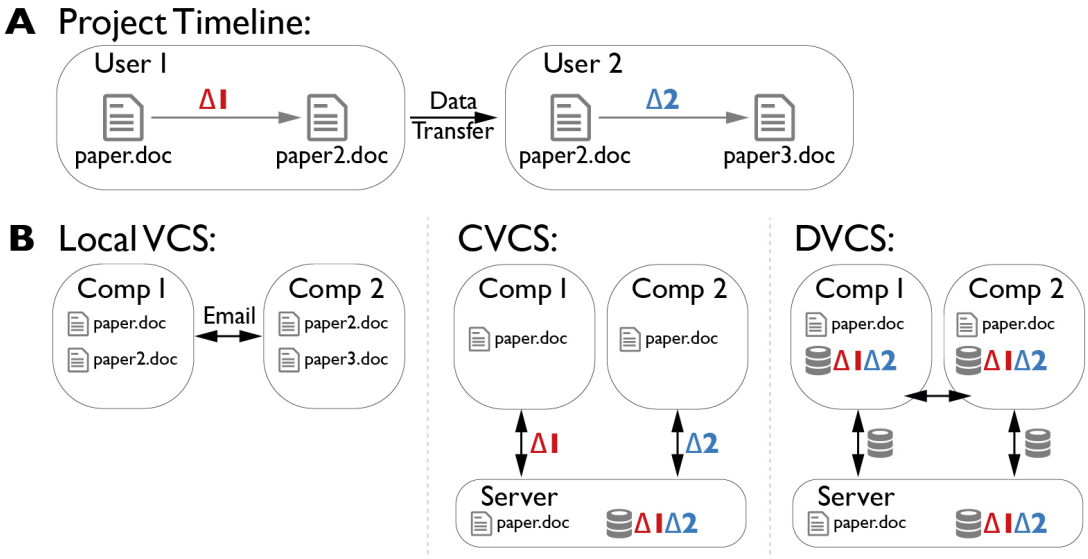


Figure 2-4. Comparison of Version Control Systems. A: An example project timeline using a rudimentary local version control system (VCS). User 1 and User 2 collaborate on a paper. User 1 makes some changes to a preexisting document 'paper.doc' and subsequently saves it as 'paper2.doc'. The set of changes by User 1 are denoted by **ΔI**. User 2 receives 'paper2.doc' from User 1 and implements their own changes, represented by **Δ2**, then saving the document as 'paper3.doc'. **B:** Infrastructure of local VCS, centralized VCS (CVCS), and distributed VCS (DVCS) using the project history

described in **A**. 1) Using a local VCS between two users resulted in many versions being loosely spread onto two computers. User 1 neither has access to, nor is aware of the most recent version 'paper3.doc'. Collaboration required the direct exchange of files from one user to another via email. 2) By introducing a central server to store the combined sets of changes from each user in a database in addition to the initial project files, a CVCS eliminates the necessity to create new files for each version on the local computers. Any version of the project can be reconstructed into a local working copy by tracing back a set of changes. 3) Avoiding the weak point of having only a single place to store the entire project history, in DVCS this information is also stored on the user's machines. This also allows offline work and a direct exchange of change sets between users. Adapted from Chacon & Straub, (2014).

Instead of only keeping a working copy of the most recent version of the project, in DVCS, all contributors keep the entire project history locally, in addition to storing it on a server. For projects with several, large, uncompressible files or with a long history of changes, this system may be limited by the amount of available hard disk space or the data transfer speed between server and clients (Lionetti, 2012). However, it excels at safeguarding against failure. When the central server of a CVCS crashes or becomes unavailable, users cannot save their changes. Unless backups of the central server are kept, the project history is lost in case a breakdown. With DVCS the project can easily be restored from any of the contributors' local copies. Furthermore, contributors don't need to be connected to a central server to save their changes, and can exchange their edits with each other before fully applying them to the project (Lionetti, 2012). This is one reason why DCVS are preferred among developers; another is that DVCS compared to CVCS allow for a higher-quality workflow, as they encourage to version changes more often (Brindescu, Codoban, Shmarkatiuk, & Dig, 2014). Git (https://git-scm.com/) and mercurial (https://www.mercurial-scm.org/) are the most common examples of DVCS.

2.3.2. Unit Testing and Test Driven Development

Yet another development practice, which improves the quality of software in addition to stringent version control, is the use of unit tests and the subsequent adoption of test driven development (TDD).

Larger programs usually consist of several functional units of code, also referred to as functions, that carry out defined tasks. Some units may be executed several times at different points of the program's overall runtime, while others may only be called once.

A unit tests is a function that is executed separately from the main program. It asserts that a corresponding programmatic function returns the expected output for a defined input (Figure 2-5). In order to be able to trace errors right back to their source, unit tests ideally check the output of each function separately. By making sure that the individual parts of a program are behaving as expected, the confidence in the program's total output remains high. Thus, rigorous unit testing helps to identify errors early on in the development process, and maintains that code behaves like expected even when functions are rewritten later (Nagappan & Maximilien, 2008).

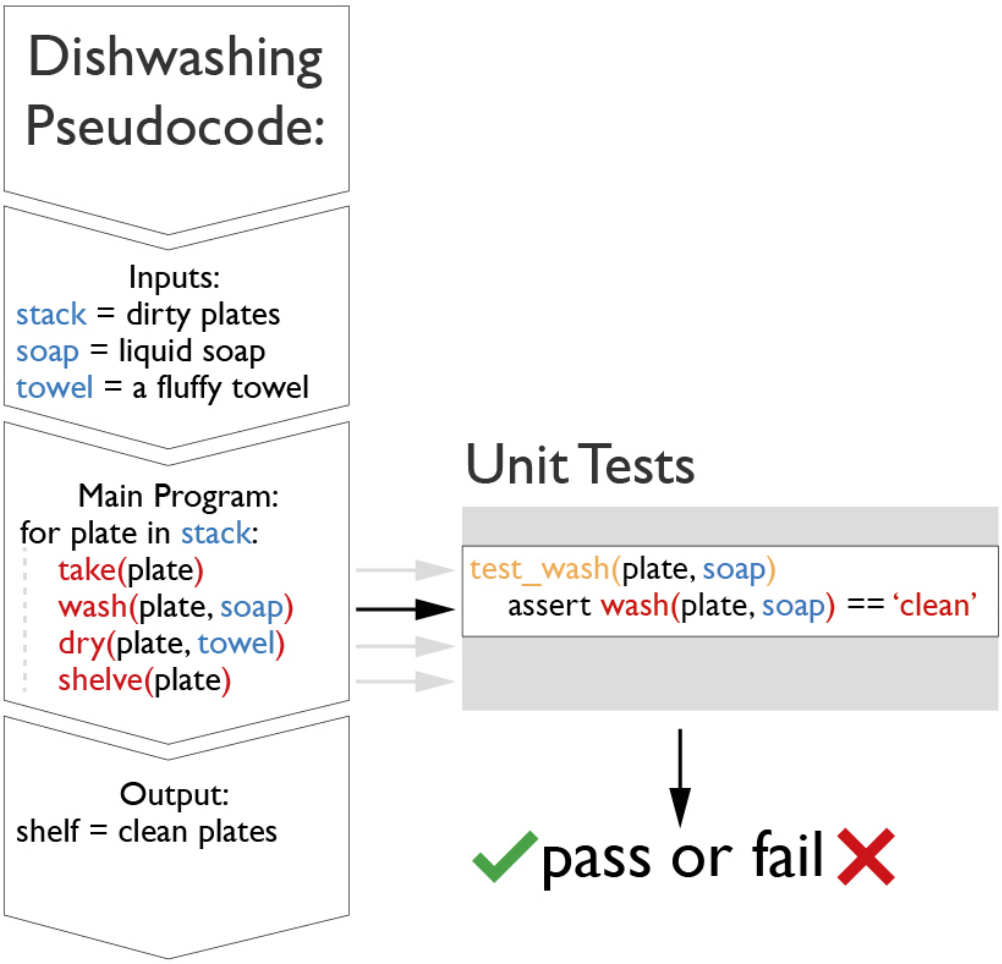


Figure 2-5. Unit tests explained with simplified pseudocode. The depicted pseudocode program for dishwashing takes a stack of dirty plates, liquid soap and a towel as inputs. The main program runs through a loop of passing each plate from the stack into sequential functions (**take()**, **wash()**, **dry()**, **shelve()**). The output is a shelf of clean plates. To facilitate debugging of the program, unit tests have been implemented for each function in isolation. For instance, **test_wash** is a test case that checks if the function **wash()** returns clean plates. It passes if this is the case, and fails if not.

The workflow of TDD requires developers to iteratively write unit tests before they write the corresponding functions that are to be tested. Although this takes longer than writing the function code first and adding unit tests later, it has the added benefit of providing constant feedback during active development. In an industrial study, the adoption of a TDD workflow has been shown to reduce the errors per thousand lines of code to up to 90% when compared to projects operating with conservative workflows (Nagappan & Maximilien, 2008).

2.3.3. Continuous Integration

Building upon version control and unit testing, continuous integration (CI) unifies a collection workflows and software tools into a larger automated process. In addition to a version control server and regardless of whether the project is maintained with CVCS or DVCS, a second dedicated CI server is set up. Using either hosted (Travis CI - https://travis-ci.org/) or self-hosted options such as Jenkins (https://jenkins.io), the CI server is configured to monitor the main project repository for changes. A great number of CI tools have been reviewed in (Shahin, Babar, & Zhu, 2017).

Whenever a developer saves a set of changes, the CI server automatically checks out that particular version of the software and executes all associated unit tests (Fowler, 2006). This happens within one or more freshly built system environments to a predefined set of instructions. For instance, a CI server can be configured to test python software with specific versions of python such as py2.7, py3.3 and py3.6, to ensure that the code is backwards compatible. The results and error logs from all builds and unit tests are then displayed in a report bearing a unique label for reference (Figure 2-6).

Figure 2-6. Example of the CI workflow using a DVCS. Continuous Integration unites version control and unit testing into an automated pipeline. A user makes incremental changes to a project, denoted by **Δ1**, commits them to the local database of changes, and finally pushes them to the DCVS server. A CI server picks up on those changes in the main project repository automatically and begins to build the testing environments, in this example, one using python2 and one using python3. Within each of those environments respectively, the CI server executes each unit test associated with the project. When the test functions have completed running the server displays a report on the status of each build.

Continuous integration principles are most beneficial when developers push and integrate their changes to the main repository often. Further, small increments

make it much simpler to identify at which point errors have been introduced. Since the entire test suite is run at each pushed change, a CI workflow inevitably enforces developers to write efficient unit tests to keep the overall runtime low (Shahin et al., 2017). In addition to guiding the team towards improving the source code, the test result statistics obtained from the CI reports, can further be used to build confidence and advertise a high-quality, trustworthy program (Meyer, 2014).

Academic research into benefits of using CI has been scarce, although CI has been widely adopted in the software industry (Shahin et al., 2017). It has, however, been shown that employing CI becomes more beneficial as the amount of contributors to a project increases, while the manpower invested into the maintenance of a CI system stays constant (Manglaviti, Coronado-montoya, Gallaba, & Mcintosh, 2017).

2.3.4. Continuous Reconstruction

The practices and tools of modern software development described above focus on automation and error prevention, while enabling distributed developers to work collaboratively. In addition to industrial monoliths such as Microsoft or IBM (Nagappan & Maximilien, 2008), many open-source projects, which build software to support scientific analyses, have implemented these approaches. This includes among countless others, Biopython (Cock et al., 2009), a python-based bioinformatics and computational molecular biology toolbox and the constraint-based reconstruction and analysis tools for MATLAB and python, COBRA toolbox (Heirendt et al., 2017) and COBRApy (Ebrahim et al., 2013) respectively.

Considering that the COBRA community has readily adopted these engineering approaches and recalling the analogy between GEM reconstruction and software development from the beginning of this sub-chapter, it is not difficult to imagine that these could be combined. With 'memote', the python tool that I introduce in Chapter 4 of this thesis, we integrate these approaches. Memote supplies a collection of unit tests for GEMs and supports the setup of a version-controlled

repository. Thus, it not only facilitates the reconstruction of GEMs in a test-driven, distributed and collaborative manner, but also functions as a standalone utility for quality control. Of course, the tool itself has been developed capitalizing on DVCS, unit testing and CI.

Chapter 3. Microbial C1 Metabolism: Recent Metabolic Modeling Efforts And Their Applications In Industrial Biotechnology (Manuscript 1)

This chapter represents a manuscript that is currently in preparation. It thus deviates from the introductory chapters in that its structure follows the typical format of a scientific paper. It has been authored by Christian Lieven, Markus J. Herrgard, and Nikolaus Sonnenschein. Christian Lieven conducted most of the work.

3.1. Abstract

To develop methanotrophic bacteria into cell factories that meet the chemical demand of the future could be both economical and ecological. Methane is not only an abundant, low cost resource, but also a potent greenhouse gas, the capture of which could help to reduce greenhouse gas emissions. Rational strain design workflows rely on the availability of carefully combined knowledge often in the form of genome-scale metabolic models to construct high-producer organisms. In this mini-review, we summarize the most recent reports on genome-scale modeling in methanotrophy and further suggest organisms that may be of interest for expanding one-carbon industrial biotechnology.

3.2. Introduction

Methane, the primary component of shale gas, natural gas and biogas, is an abundant, albeit highly distributed and small-scale resource (Clomburg et al., 2017). A powerful greenhouse gas, its release from decomposing landfill and agricultural waste, gas flares, and wastewater treatment plants into the atmosphere contributes strongly to global warming (Bennett et al., 2018). The output from these sites can be captured and the carbon, that is currently wasted, can sustainably be converted into value-added chemicals, fuels or electricity by means

of microbial activity. To illustrate, the amount of carbon released from global venting and flaring in 2014 alone would have been sufficient to cover the world's requirement for methanol, ethylene, propylene, butadiene, xylene, benzene, and toluene (Clomburg et al., 2017). In addition to the environmental benefits of carbon capture at these sites, the price of methane is lower and its per-carbon yield higher than that of glucose, making it the ideal substrate for cell factories (Comer, Long, Reed, & Pfleger, 2017). Consequently, public and private funding, and thus general research in this area has increased.

A number of studies that explore potential gas-to-products technologies using methanotrophic organisms have been reviewed by Pieja, Morse, & Cal (2017). Methanol, which is the first product of aerobic methane oxidation, presents a similarly suitable feedstock for biotechnological applications. Strategies involving native methylotrophs have been reviewed by Clomburg et al. (2017), while achievements in synthetic implementations of methylotrophy have been expanded upon by Bennett et al. (2018). To rationally improve strain designs of native and synthetic methanotrophs, *in silico* systems biology tools can be employed (Ng, Khodayari, Chowdhury, & Maranas, 2015). The fundament of many *in silico* approaches is a formalized representation of an organism's metabolic network in the form of a genome-scale metabolic model (GEM).

Here in this minireview, we focus on organisms for which GEMs are currently available in literature that could support the development and improvement of industrial producer strains, which convert methane or methanol into value-added compounds. Elsewhere, a similar effort has been carried out investigating genome-scale metabolic models of clostridia, which are the relevant biocatalyst of syngas (CO_2, CO, H_2) fermentations (Dash, Ng, & Maranas, 2016).

3.3. Underlying principles of genome-scale metabolic modeling

Constraints-based reconstruction and analysis (COBRA) of metabolic networks has become a widely adopted discipline of systems biology in the past two decades (Palson, 2015). From the sequenced genome of any given organism, information about the specific enzymatic reactions can be extracted and translated into a set of stoichiometric equations. These equations are then viewed as a closed system, which is mass-, and ideally, charge-balanced. Moreover, it is assumed that the system is at a steady state, meaning that there is no net accumulation of intracellular material. Based on these premises, the internal metabolic fluxes of an organism can be expressed as: $\mathbf{S}\,\mathbf{v} = \mathbf{0}$, a linear system of equations where the matrix $\mathbf{S}$ represents all stoichiometric coefficients from the set of enzymatic reactions, and the vector $\mathbf{v}$ represents the flux distribution across all reactions. Flux balance analysis (FBA) can then be used to obtain a specific flux distribution, typically one that maximizes the flux through a specific reaction, for instance the biomass equation (Cotten & Reed, 2016).

A well-curated genome-scale metabolic network by itself is a powerful knowledgebase as it accounts for the interconnection between genes, reactions, metabolites and meta-information (Thiele & Palsson, 2010). Built on top of this, FBA and derivative methods have been shown to accurately predict growth phenotypes. Thus, they are useful to prospect strategies for metabolic engineering *in silico* (Erb, Jones, & Bar-Even, 2017). One great example for the successful integration of a genome-scale metabolic model and metabolic engineering is the development of an *E. coli* strain capable of producing 1,4-butanediol (Yim et al., 2011), the design of which relied on the available GEM at the time (Reed, Vo, Schilling, & Palsson, 2003) and an algorithm for the prediction of biological pathways to a specific target compound (Cho, Yun, Park, Lee, & Park, 2010). The strain has been patented and since 2013 has been licensed by BASF (BASF, 2013). Since then many more fruitful applications of GEMs and COBRA methods have

been reported, although chiefly for the model organisms *Escherichia coli* and *Saccharomyces cerevisiae* (Simeonidis & Price, 2015).

3.4. State of the art: Existing C1 metabolic models.

3.4.1. GEMs for Aerobe Methanotrophy

Nature has found two distinct ways of breaking the strong bond between the one carbon atom and one of the four hydrogen atoms of methane (Bollinger & Broderick, 2009). In aerobic methanotrophs, two types of methane-monooxygenases can catalyze this reaction, converting methane and oxygen to methanol and water. Few organisms express a soluble monooxygenase (sMMO), which receives the electrons necessary for oxidation from NADH. In a majority of methanotrophs, however, the reaction is carried out by a membrane-bound, so called particulate methane-monooxygenase (pMMO). While it also requires two electrons to carry out the oxidation of methane, its native reductant is still debated (Lawton & Rosenzweig, 2016a). Three possible scenarios regarding the mode of electron transfer can be considered (Figure 3-1): 1) electrons needed for the oxidation of methane are supplied by NADH produced from formaldehyde oxidation further downstream, while the electrons from methanol oxidation are shuttled into ATP production via a *redox arm* composed of cytochromes. 2) The pMMO is *directly coupled* to the methanol dehydrogenase, which allows an immediate exchange of electrons between the two reactions. 3) In the so called *uphill electron transfer* electrons are supplied to the pMMO by MDH and NADH (Leak & Dalton, 1986b).

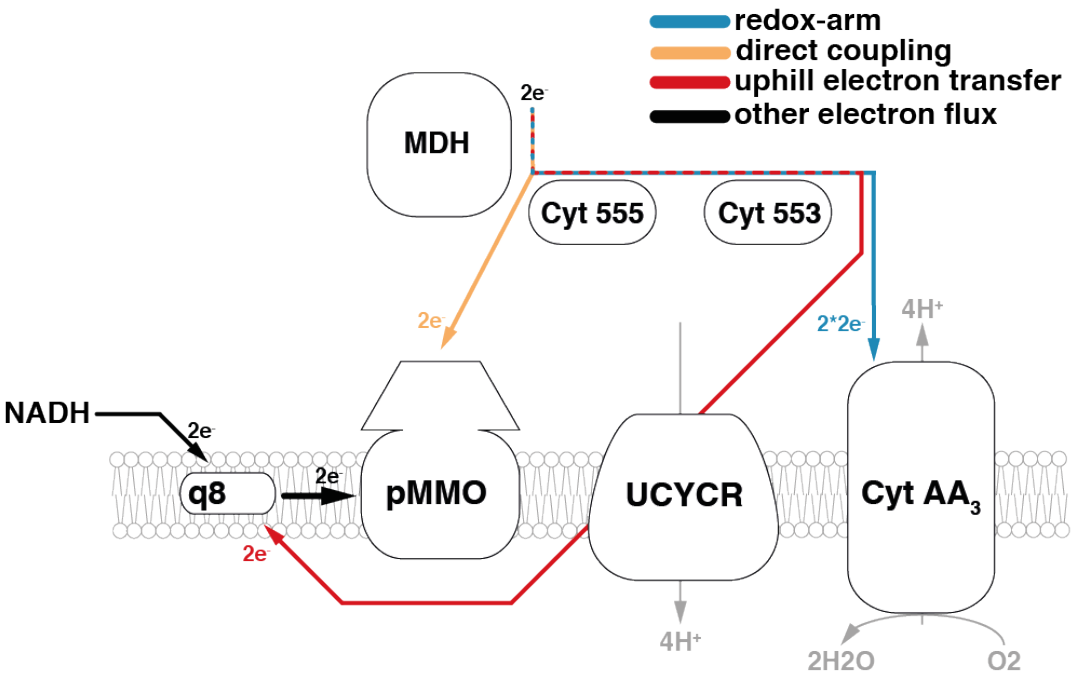


Figure 3-1 The three putative modes of electron transfer to the pMMO. MDH = methanol dehydrogenase, Cyt 555 = cytochrome c555, Cyt 553 = cytochrome c553, Cyt AA_3 = terminal oxidase, UCYCR = ubiquinol-cytochrome-c reductase, pMMO = particulate methane monooxygenase, q8 = ubiquinol pool

With the genome-scale metabolic model of *Methylomicrobium buryatense* 5G(B1), de la Torre et al. (2015) presented the first ever manually curated GEM of a methane-utilizing bacterium. Using the model to explore each mode of electron transfer, they were able to eliminate the *redox arm* hypothesis since it correlated the least with their experimental measurements for *M. buryatense*. Further the researchers found that the *in silico* replacement of the pyruvate dehydrogenase with a phosphoketolase did not improve the overall carbon yield on methane. Instead a decrease in carbon yield was predicted since more methane was oxidized to CO2 as a source of NADH. The genome scale metabolic model complements the genetic tools that have already been established for *M. buryatense* (Puri et al., 2015), and recent successes in metabolic engineering show that there is potential to commercialize production in *M. buryatense*. For instance, Dong et al. (2017) have been able to increase the production of membrane phospholipids, which they subsequently processed into diesel blendstock, and Henard et al. (2016) have used *M. buryatense* to produce lactate at a yield of 0.05 g lactate/g methane. In addition,

the latter have been able to improve the lipid and biomass yield 2.6-fold by overexpressing the phosphoketolase (Henard, Smith, & Guarnieri, 2017).

A GEM of the halotolerant *Methylomicrobium alcaliphilum* 20ZR is also available from the Kalyuzhnaya Lab (Akberdin et al., 2018). The authors gathered metabolomics profiles of *M. alcaliphilum* grown on methane and methanol to verify and improve the *in silico* predictions. The model correctly predicted an increase in the metabolite pools of amino acids when grown on methanol instead of methane. Furthermore, simulations indicated that during oxygen-limited growth pyrophosphate-dependent reactions play an important role to improve the biomass yield. Lastly, they determined that a *direct coupling* electron transfer mode fit best their observations (Figure 3-1). In previous studies, the organism was found to ferment methane-derived formaldehyde to organic acids at low oxygen tension (Kalyuzhnaya et al., 2013), and was used to investigate the biosynthesis and degradation pathways of sucrose in methanotrophs (But et al., 2015). Hill, Chrisler, Beliaev, & Bernstein (2017) successfully co-cultivated *M. alcaliphilum* 20Z with *Synechococcus* PCC 7002 on a mixture of methane and carbon dioxide. The potential for the production of the osmolyte and biostabilizer ectoine in *M. alcaliphilum* 20Z may, however, hold promise for an economically viable industrial application. One kilogram of ectoine costs around US$ 1000, with the average global demand of the pharmaceutical industry amounting to 15,000 tonnes per year (Strong, Xie, & Clarke, 2015). Similar to the simulations carried out by Ates, Oner, & Arga (2011) for *Chromohalobacter salexigens* DSM 3043, the GEM for *M. alcaliphilum* 20ZR could be employed to prospect ectoine production and develop hypothesis-driven strain engineering strategies.

The production of animal feed from natural gas using *Methylococcus capsulatus* as the provider of single cell protein (SCP) had already been commercialized in the 70s until a drop in oil prices made these efforts economically infeasible. Today, two companies produce SCP at pilot scale, US-based Calysta (www.calysta.com) and Denmark-based Unibio (www.unibio.dk). In a joint venture with the global conglomerate Cargill, Calysta is constructing a commercial scale plant with an

expected production capacity of 200,000 tonnes per year. The estimated date of completion is 2020. In addition, several patents have been filed for the production of chemicals in *M. capsulatus*: propylene by Calysta (Silverman & Purcell, 2014), succinate by String Bio (Subbian, 2017), 1,4-butanediol by Sekisui Chemical (Furutani, Uenishi, & Iwasa, 2015) and other multi-carbon compounds by the Intrexon Corporation (Coleman et al., 2017). Although *M. capsulatus* has been extensively studied in the past 50 years (Trotsenko & Murrell, 2008), a curated GEM was only recently completed by Lieven, Gernaey, Herrgard, & Sonnenschein (2018). Here, the authors computationally predicted transporter genes and assigned them to corresponding transport reactions. Similar to the efforts carried out by de la Torre et al. (2015) for *M. buryatense*, they investigated which mode of electron transfer best represents measured parameters for *Methylococcus capsulatus*. They found that simulations of the three modes exclusively could adequately represent the experimentally observed ratio of O2 uptake per mol of methane. Only by reducing the efficiency of the *uphill electron transfer* mode were they able to replicate the reference ratio. Moreover, they found, that the energetic burden of NH_4 oxidation to NO_2 by the pMMO likely affects this ratio, when cells are grown on medium containing NH_4 as the nitrogen source. To facilitate visual inspection of multi-omics data and more intuitive exploration of the metabolic potential, the authors also provide a metabolic map that displays the metabolic network described by the model.

3.4.2. GEMs for Methylotrophy

Despite being the more reduced, the hypothetical per carbon substrate yields of methane have been determined to be consistently lower than those of methanol in an *in silico* study carried out by Comer et al. (2017). This is due to the low efficiency conversion catalyzed by the methane monooxygenase. The natural, aerobic methane oxidation requires two electrons, which subsequently have to be recovered by the methanol dehydrogenase, oxidizing methanol to formaldehyde, or further downstream depending on the mode of electron transfer (Clomburg et al., 2017). This contributes to loosing 36% of the energy within the highly reduced molecule by an essentially redox-neutral conversion of methane to formaldehyde

(Haynes & Gonzalez, 2014). Since methanol is an intermediate of methanotrophy many methanotrophs can use it as their sole carbon and energy source. Since it is a liquid, using methanol bypasses potential issues with mass-transfer during gas-fermentation. Thus, methanol represents a potential alternative single-carbon feedstock.

The metabolism of the facultative methylotroph *Methylobacterium extorquens* AM1 has been studied in detail since well over 50 years (Christopher Anthony, 2011). The considerable research interest has resulted in the development of genetic tools and protocols, ultimately leading to the establishment of several production processes, the products of which include polyhydroxyalkanoates (PHA), serine, dicarboxylic acids derived from the ethylmalonyl-CoA pathway, alcohols and proteins (Ochsner, Sonntag, Buchhaupt, Schrader, & Vorholt, 2014). A GEM was established for *M. extorquens* AM1 to investigate the topology and operation of the intertwined metabolic cycles that operate in the bacterium when grown on methanol (Peyraud et al., 2011). In a separate study, researchers succeeded in heterologously producing 1.65 g/L of α-homulene, an anti-inflammatory terpenoid, in *M. extorquens*. The metabolic model was used to calculate the maximum theoretical yield of the compound.

3.4.3. GEMs for Anaerobic Methanotrophy

In addition to the energy loss caused by the oxygen-dependent conversion of methane mentioned above, the volumetric mass transfer of methane and oxygen is another limitation especially at large-scale operation. Although innovative specialized reactor designs alleviate the issue (Petersen, Villadsen, Jørgensen, & Gernaey, 2017), they translate into increased capital expenses when compared to using regular stirred-tank vessels. With respect to these drawbacks, the anaerobic production of chemicals from methane is considered more ideal, despite it exhibiting lower growth rates and productivity (Bennett et al., 2018).

The methyl-coenzyme M reductase (MCR) is the key enzyme required for the anaerobic biosynthesis of methane. In this reaction, methyl-coenzyme M (methyl-

SCoM) reacts with coenzyme B (CoBSH) to form methane and COBS-SCoM (Lawton & Rosenzweig, 2016a). Anaerobic methanotrophs (ANME) avail themselves of a homolog MCR that is able to catalyze the reverse reaction. In nature, ANME grow in consortia with syntrophic bacteria that participate in the removal of reducing equivalents, which has complicated the isolation and culturing of native ANME strains (Bennett et al., 2018). However, through metagenomic sequencing it was recently possible to obtain the corresponding MCR gene, successfully clone, and express it in the methane-producing archaeon *Methanosarcina acetivorans* C2A (Soo et al., 2016). The authors updated the two existing GEMs for *M. acetivorans,* iVS941 (Satish Kumar, Ferry, & Maranas, 2011) and iMB754 (Benedict, Gonnerman, Metcalf, & Price, 2012), to study the feasibility of producing acetate, formate and pyruvate on methane as a function of Fe^{3+} reduction. This updated GEM, named iMAC868, was then published by Nazem-Bokaee, Gopalakrishnan, Ferry, Wood, & Maranas (2016). With iMAC868 they improved the predictions of growth yield on the native substrates methanol and acetate, in addition to making the necessary changes to enable methanotrophy. The authors predicted the hypothetically maximal biomass yields and the yields of biofuel precursors methanol, ethanol, butanol and isobutanol on methane. Considering the ΔG of different external electron acceptors, they found that the yields were highest for Fe^{3+} reduction when CO2 in the form of bicarbonate was co-utilized. The native products of reverse methanogenesis in *A. acetivorans* were determined to be acetate and CO2. Using the same engineered host, McAnulty et al. (2017) produced lactate yielding 0.59 g per gram of methane. This is an order of magnitude greater than the previously mentioned yield of lactate on methane in an aerobic process (Henard et al., 2016).

Bennett et al. (2018) suggest that since there are genetic tools available for it, the hydrogenotrophic methanogen *Methanococcus maripaludis* could also be consider as a host for reverse methanogenesis. Several metabolic models currently have been reconstructed for this archaeon with the most recent one being iMR539 (Richards et al., 2016).

3.4.4. Approaches for Synthetic Methanotrophy

Slow growth rates, inefficient molecular techniques and a lack of experience compared to model microorganisms complicate the work with native aerobic and anaerobic methanotrophs. While the implementation of aerobic methane oxidation using pMMO or sMMO has been difficult (Balasubramanian et al., 2010; Gou et al., 2006), the transfer of precursor pathways belonging to anaerobic methanotrophy has been successful (Scheller, Yu, Chadwick, McGlynn, & Orphan, 2016), thus making the prospect of synthetic anaerobic methanotrophy more promising. Most progress, however, has been made with the heterologous expression of methylotrophic genes in the microbial workhorses *Escherichia coli, Corynebacterium glutamicum* and *Saccharomyces cerevisiae* (Gonzalez, Bennett, Papoutsakis, & Antoniewicz, 2018; Nguyen, Hwang, Chan, & Lee, 2016; Whitaker et al., 2017). It is no surprise that GEMs for these three well-established model organisms exist and are continuously updated. For reference, the most recent versions are listed in Table 3-1.

Another promising option, which can be regarded as a step towards heterologous methanotrophy, is the development of synthetic pathways involving novel enzymes (Erb et al., 2017). The formolase pathway, which has been constructed around the computationally designed enzyme formolase (FLS), is such a pathway. Overall, the five-step, linear pathway catalyzes the carboligation of three formate molecules into the common three-carbon intermediate dihydroxyacetylphosphate (DHAP) and has been shown to function *in vitro* (Siegel et al., 2015). The authors used flux balance analysis and the core metabolic model of *E. coli* (Orth, Palsson, & Fleming, 2010) to compare the performance of their novel pathway relative to all natural formate assimilation pathways. They found that the hypothetical maximum biomass yield of the formolase pathway is the second highest (6.5 g cell dry weight/mol formate) behind the reductive TCA cycle (6.7 g cell dry weight/mol formate), but exceeds all other pathways when considering the chemical driving force.

Table 3-1. Genome-scale metabolic models relevant to methano- or methylotrophy.

Organism	Model ID	Previous Versions	# Reactions	# Metabolites	# Genes	Reference
Aerobic						
Methylobacterium extorquens AM1	iRP911[a]	First	1139	977	911	Peyraud et al., 2011
Methylococcus capsulatus Bath	iCL730 [a, c]	First	898	877	730	Lieven et al., 2018
Methylomicrobium alcaliphilum 20Z	iIA407 [b]	First	433	423	407	Akberdin et al., 2018
Methylomicrobium buryatense 5G(B1)	i5GB1[a, b]	First	402	403	314	de la Torre et al., 2015
Anaerobic (Chassis for Reverse Methanogenesis)						
Methanococcus maripaludis S2	iMM518[b]	First	570	556	518	Goyal, Widiastuti, Karimi, & Zhou, 2014
Methanococcus maripaludis S2	iMR539[a, c]	Independent from iMM518	688	710	539	Richards et al., 2016
Methanosarcina acetivorans C2A	iMAC868[b]	iVS941[d], iMB745[e]	845	718	868	Nazem-Bokaee et al., 2016
Sythetic Methano- or Methylotrophy						
Corynebacterium glutamicum	iCW773[b]	Reviewed by Milne, Kim, Eddy, & Price, 2009	1207	950	773	Zhang et al., 2017
Eschericia coli	iML1515[a]	Reviewed by McCloskey, Palsson, & Feist, 2013	2712	1877	1516	J. M. Monk et al., 2017
Sacharromyces cerevisiae	YEAST 7[a, c]	Reviewed by Heavner & Price, 2015	3493	2220	909	Aung, Henry, & Walker, 2013

[a]Available as SBML, [b]Available as XML, [c]Available as MAT, [d]Satish Kumar et al., 2011, [e]Benedict et al., 2012

3.5. Conclusion

A well-curated genome-scale metabolic model is particularly useful for exploring the topology and systems properties of metabolism. In addition, a GEM establishes a connection between stoichiometric, genetic, and meta information which is the foundation of many strain-design methods. As evident from this review, metabolic models of methanotrophy are scarce.

Many more organisms than the ones reviewed here could become relevant as methanotrophic producer strains, and thus could benefit from having a GEM available. Yet, the underlying requirement for a high-quality GEM is the existence of an expertly annotated genome sequence. Collected in Table 3-2 are potential organisms of interest, and references to the corresponding genome sequences.

In spite of their potential as producer organisms being unclear, having access to curated biochemical information can still benefit the scientific community as a whole. As J. Monk et al. (2014) remark, the coverage of metabolic reactomes has stagnated, since little effort is spent on comprehensively uncovering the metabolic space of an organism, especially with regards to the secondary metabolism. A thorough analysis of these organisms may lead to interesting discoveries similar to, for instance, that of hopanoid production in *M. capsulatus* and *Alicyclobacillus acidocaldarius* (Belin et al., 2018).

Primarily using flux balance analysis, the metabolic models reported here have been used to explore metabolic interconnections and the system's behavior in specific conditions. To improve the predictiveness of GEMs, however, a cross validation with genetic perturbation experiments in addition to growth studies is an invaluable step (Machado et al., 2016; Thiele & Palsson, 2010).

Table 3-2. Methanotrophs that are potentially relevant as biotechnological producers. PHB = polyhydroxybutyrate

Organism	Genome Sequence	Primary Interest
Methylobacter marinus 7C	Flynn et al., 2016	Identification of the ectoine biosynthesis genes. (Reshetnikov et al., 2011)
Methylobacterium organophilum CZ-2	Not sequenced	Production of PHB (Zúñiga, Morales, Le Borgne, & Revah, 2011) and triacylglycerides (Strong et al., 2015).
Methylocaldum sp. SAD2	Not sequenced	Production of methanol on high levels of H_2S (W. Zhang, Ge, Li, Yu, & Li, 2016).
Methylocapsa acidiphila	Direct submission NZ_ATYA00000000.1	Potential production of PHB (Dedysh et al., 2018).
Methylocella tundrae	Not sequenced	Production of methanol (Mardina et al., 2016).
Methylocystis bryophila	Direct submission NZ_CP019948.1	Production of methanol (Patel, Mardina, Sang-Young, Jung-Kul, & In-Won, 2016).
Methylocystis parvus OBBP	del Cerro et al., 2012	Production of PHB (Rostkowski, Pfluger, & Criddle, 2013)
Methylocystis sp. WRRC1	Not sequenced	Production of a copolymer of PHB and hydroxyvalerate(Cal et al., 2016).
Methylomicrobium kenyense AMO1	Not sequenced	Identification of the ectoine biosynthesis genes. (Reshetnikov et al., 2011)
Methylomonas denitrificans	Kits, Klotz, & Stein, 2015	Production of N_2O coupled to methane oxidation under hypoxia (Kits et al., 2015).
Methylomonas sp. 16a	Sequenced by DuPont, unpublished	Synthesis of C30 carotenoids (Tao, Schenzle, & Odom, 2005), production of astaxanthin and canthaxanthin (Sharpe et al., 2007; Tao et al., 2007).
Methylosinus sporium	Not sequenced	Production of methanol (Patel, Selvaraj, et al., 2016).
Methylosinus trichosporium OB3b	Stein et al., 2010	Production of methanol (Ge, Yang, Sheets, Yu, & Li, 2014)

Promiscuous enzyme functions are often not included in biochemical databases and rarely covered in the genome annotation (Fiehn, Barupal, & Kind, 2011). While the methanogen models iMM518 (Goyal, Widiastuti, Karimi, & Zhou Zhi, 2013), iMR539 (Richards et al., 2016) and iMAC868 (Nazem-Bokaee et al., 2016) have been validated using small-scale knockout data, the use of this method of validation for a GEM of a methanotroph has been limited to a single reaction knockout in iIA407 (Akberdin et al., 2018). Yet, Richards et al. (2016) point out that this is made difficult by a low abundance of suitable gene knockout data.

The application of automated strain design methods in this field has not been reported so far. A host of strain design methods have been thoroughly reviewed by Ng et al. (2015). Applying a pathway prediction method such as GEM-Path, for instance, could decrease the time required to create a suitable design for the production of commodity chemicals from methane. The algorithm provided 1271 growth-coupled designs for the production of 20 commodity chemicals in *E. coli* (Campodonico, Andrews, Asenjo, Palsson, & Feist, 2014). Using the metabolic models that are currently available for methanotrophs, the production potential of the organisms could already be explored.

As evident from Table 3-1, many GEMs are distributed in a non-standard, tabular file format. While the MATLAB version of the COBRA Toolbox is able to import and simulate models that come in this format, many other software tools rely on the communication of models in the Systems Biology Markup Language (SBML). In fact, Ravikrishnan & Raman (2015) advocate the distribution of models in this *de facto* community standard, because the use of other formats may decrease reproducibility and the ability to use a GEM with the largest portion of available tools (Ravikrishnan & Raman, 2015).

Although the true impact of GEMs in a specific field of research is hard to quantify, their general success is indisputable (Kim, Kim, & Lee, 2017). Along with this, and as a result from the reinvigorated interest in methanotrophy, we are able to summarize the 4 available GEMs for aerobic methyl- and methanotrophy in this work. Despite the minor shortcomings noted above, the availability of

these models can only serve to accelerate the process of discovery. Allowing researchers to probe certain properties *in silico*, isolated from the potential challenges associated with slow-growth, difficulties to culture an organism, inefficiencies of molecular techniques or even the lack thereof, GEMs represent an ideal tool to explore and expand the metabolic potential of methanotrophs. The first steps, presented here, may lead towards a solid foundation for rational strain-design of methanotrophs, and may help to elucidate their many puzzles.

3.6. Acknowledgements

The authors acknowledge financial support from Innovation Fund Denmark (project "Environmentally Friendly Protein Production (EFPro2)") and the Novo Nordisk Foundation.

Chapter 4. Memote: A community driven effort towards a standardized genome-scale metabolic model test suite (Manuscript 2)

This chapter represents a manuscript that is currently in preparation. It deviates from the previous chapter in that its structure follows the requirements for the submission type of a Resource at Nature Biotechnology. It has been authored by Christian Lieven, Moritz E. Beber, Brett G. Olivier, Frank T. Bergmann, Parizad Babaei, Andreas Dräger, Janaka N. Edirisinghe, Ronan M. T. Fleming, Beatriz García-Jiménez, Wout van Helvoirt, Henning Hermjakob, Markus J. Herrgård, Hyun Uk Kim, Zachary King, Jasper J. Koehorst, Steffen Klamt, Meiyappan Lakshmanan, Nicolas Le Novère, Dong-Yup Lee, Sang Yup Lee, Sunjae Lee, Nathan E. Lewis, Daniel Machado, Adil Mardinoglu, Gregory L. Medlock, Jonathan M. Monk, Jens Nielsen, Juan Nogales, Intawat Nookaew, Osbaldo Resendis-Antonio, Bernhard Ø. Palsson, Jason A. Papin, Kiran R. Patil, Nathan D. Price, Anne Richelle, Peter J. Schaap, Rahuman S. Malik Sheriff, Saeed Shoaie, Nikolaus Sonnenschein, Bas Teusink, Ines Thiele, Jon Olav Vik, Joana Xavier, Maksim Zakhartsev, and Cheng Zhang. Christian Lieven wrote all parts of the manuscript except for Chapter 4.3.1, which was written by Brett G. Olivier and Frank T. Bergmann. Moritz E. Beber and Christian further devised and implemented the software.

4.1. Abstract:

Several studies have shown that neither the formal representation nor the functional requirements of genome-scale metabolic models (GEMs) are clearly defined. Without a consistent standard, comparability, reproducibility and interoperability of models across groups and software tools cannot be guaranteed.

Here, we present memote, an open-source software containing a community-maintained, standardized set of metabolic model tests. The tests cover a range of

aspects from annotations to conceptual integrity, and can be extended to include experimental datasets for automated model validation. In addition to testing a model once, memote can be configured to do so automatically i.e. during GEM reconstruction. A comprehensive report displays the model's performance parameters, which supports informed model development and facilitates error detection.

Memote provides a measure for model quality that is consistent across reconstruction platforms and analysis software, and simplifies collaboration within the community by establishing workflows for publicly hosted and version controlled models.

4.2. Main

The reconstruction and analysis of metabolic reaction networks provides mechanistic, testable hypotheses for an organism's metabolism under a wide range of empirical conditions (Palsson 2015). At the current state of the art, genome-scale metabolic models (GEMs) can include up to thousands of metabolites and reactions assigned to subcellular locations, gene-protein-reaction rules (GPR), and annotations which provide meta-information by referencing large biochemical databases. This development has been facilitated by standard protocols for reconstruction (Thiele and Palsson 2010) and guidelines for provenance tracking and interoperability (Heavner and Price 2015; Wilkinson et al. 2016; van Dam et al. 2017). However, the quality control of GEMs remains a formidable challenge that must be solved to enable confident use, reuse and improvement.

Finding that GEMs can be published as SBML (Hucka et al. 2010), MATLAB files, spreadsheets, and PDF, both Ravikrishnan & Raman (2015) and Ebrahim et al. (2015) lamented the lack of an agreed-upon description format. While the former noted that incompatible formats limit the scientific exchange and thus the ability to reproduce calculations on different setups, the latter elaborated how

formatting errors can directly cause inconsistent results when parsed and evaluated with different software packages.

When comparing four previously published models for *Pseudomonas putida* KT2440 (Yuan et al., 2017), the authors discovered that in identical simulation conditions the predicted growth rate of one model was almost twice as high as that of another. Moreover, one of the examined models could generate ATP without needing to consume any substrate, rendering the model predictions useless.

This behaviour occurs when a model's reactions are not checked for thermodynamic feasibility, leading to the formation of flux cycles which provide reduced metabolites to the model without requiring nutrient uptake. Fritzemeier et al. (2017) detected such erroneous energy-generating cycles (EGCs) in the majority of GEMs specifically in the MetaNetX (Ganter et al. 2013; Moretti et al. 2016) (~66%) and ModelSEED (Henry et al. 2010) (~95%) databases, which mostly contain automatically generated, non-curated metabolic models. Although the authors found that EGCs are rare in manually curated GEMs from the BiGG Models database (~4%), their effect on the predicted growth rate in FBA may account for an increase of up to 25%. This makes studies involving the growth rates predicted from such models unreliable. It is possible to identify and correct these issues either with functions included in the COBRA Toolbox (Schellenberger et al. 2011), or the modified GlobalFit algorithm (Hartleb et al. 2016) presented by Fritzemeier et al. (2017). Yet, as the above-mentioned model of *P. putida* from Yuan et al. (2017) shows, this is not done consistently.

Investigating the biomass compositions (BCs) of 71 manually curated prokaryotic GEMs, Xavier, Patil, & Rocha (2017) found that organic cofactors (e.g., Coenzyme A, pyridoxal 5-phosphate, and S-adenosyl-methionine) are missing even though their inclusion is vital to a model's performance in gene-essentiality studies. According to Xavier et al., all models for *Mycobacterium tuberculosis* lack the ability to produce pyridoxal 5-phosphate, in spite of experimental support for its importance for growth, survival and virulence (Dick, Manjunatha, Kappes, & Gengenbacher, 2010).

Chan et al. (Chan, Cai, Wang, Simons-Senftle, & Maranas, 2017) highlighted deviations in molecular weight as another problem with the formulation of BCs. Conforming to the defined molecular weight of 1 g/mmol is essential to reliably calculate growth yields, cross-compare models, and obtain valid predictions when simulating microbial consortia. Half of the 64 tested models deviated from the defined 1 g by up to 5%, with the other half deviating even more strongly. Any discrepancy, however, should be avoided as the smallest error affects the predicted biomass yield, favouring models containing BCs which sum to a lower molecular weight.

In addition to discussing encoding related problems, Ravikrishnan & Raman (2015) stressed that missing metabolite and reaction annotations are another fundamental issue when trying to exchange GEMs, which have been generated from different platforms, or when trying to integrate them into existing computational workflows. They reported the absence of metabolite annotations (i.e., metabolite formula, database-dependent (e.g., ChEBI ID), and database-independent (e.g., SMILES, InCHI) references) in almost 60% of the 99 models they examined.

Increasing numbers of manually-curated and automatically-generated GEMs are published each year, growing both in scale and scope; from models on single cells to multi-organism communities (Kim et al. 2017) to multi-compartmental plant (Zakhartsev et al. 2016), human and cancer tissue models (Jerby and Ruppin 2014). Especially when considering the growing application of models to human health and disease it becomes essential to address any remaining issues concerning reproducibility and interoperability in order to pave the way for reliable systems medicine (Olivier et al. 2016).

Thus, we need to establish a standard framework which ensures that:

1. Models are formulated consistently in a tool independent manner.

2. Components of GEMs are uniquely identifiable using standardized database identifiers that can be converted easily using cross-references.
3. Default conditions are clearly defined to allow the reproduction of the original model predictions.
4. Models yield biologically feasible phenotypes when analysed under alternate conditions.
5. Data that has been used to constrain/parametrize the model is documented properly in order to precisely understand the model refinement process.

Here, we argue for a two-pronged approach in creating this framework. We advocate the use of the latest version of the *SBML Level 3 Flux Balance Constraints (FBC) Package* (Olivier and Bergmann 2015) as the agreed-upon description format, which renders GEMs to be independent through a unified formulation; and, borrowing tools and best practices from software development (Cooper et al. 2015; Beaulieu-Jones and Greene 2017), we present *memote* as a unified approach for benchmarking metabolic models.

4.3. Results

4.3.1. SBML: Tool independent model formulation

Historically, GEMs have been structured and stored many non-standard ways, for example, tool specific formats or language dialects (Ravikrishnan & Raman, 2015). This prevented the accurate exchange of models between various software tools and the unambiguous, machine-readable description of all model elements such as chemical reactions, metabolites, gene associations, annotations, objective functions and flux capacity constraints. While a widely used model description standard, such as the Systems Biology Markup Language (SBML) Level 3 Core (Hucka et al., 2010), can describe some of these components e.g., reactions, metabolites and annotations, it cannot present other model components needed to describe a parameterised GEM or FBA model in a structured and semantic way.

This translates into the need for an effective model description format that allows for the unambiguous definition and annotation of such a model's components and underlying mathematics.

Fortunately, such a format does exist, the *SBML Level 3 Flux Balance Constraints (FBC) Package*. Amongst other things, the FBC package allows for the unambiguous, tool neutral and validatable SBML description of flux bounds, multiple linear objective functions, gene-protein-reaction associations, metabolite chemical formulas, charge and related annotation (SBML3FBCSPEC)(Brett G. Olivier & Bergmann, 2015). The package is developed in collaboration by the SBML and constraint-based modelling communities, which has led to FBC Version 2 moving towards becoming the *de facto* standard for encoding GEMs. Critical to this process is its implementation in a wide range of constraint-based modelling software and adoption by public model repositories (Bergmann and Sauro 2006; Bornstein et al. 2008; Frank T. Bergmann 2011; Ebrahim et al. 2013; Chelliah et al. 2015; Olivier et al. 2016; King et al. 2016; Rodriguez et al. 2015; Olivier 2011).

4.3.2. Memote: Community-driven quality control

In software engineering, test-driven development ensures that in response to a defined input a piece of code generates the expected output (Nagappan, Michael Maximilien, Bhat, & Williams, 2008), distributed version control represents an efficient way of tracking and merging changes made by a group of people working on the same project (Brindescu, Codoban, Shmarkatiuk, & Dig, 2014), and continuous integration ties these two principles together by automatically triggering tests to be executed after each change that is introduced to the project (Meyer 2014). Memote (/'mi:moʊt/ (IPA)), short for metabolic model tests, is an open-source python software that applies these engineering principles to genome-scale metabolic models.

Memote allows users to benchmark stoichiometric models encoded in SBML3FBC against a set of consensus tests. By enabling researchers to quickly

interrogate the quality of GEMs, problems can be addressed before they affect reproducibility and scientific discourse, or increase the amount of time spent troubleshooting (Baker, 2016).

Memote supports two fundamental workflows (Figure 4-1a). First, by running the test suite on a model once, memote generates a comprehensive, human-readable report, that quantifies the model's performance. On the basis of this information, a definitive assessment of model quality can be made i.e. by editors or reviewers. This workflow is accessible through a web interface, analogous to the SBML validator (Bornstein et al. 2008), or locally through the command line.

Second, for model maintenance and reconstruction, memote coordinates version control and continuous integration, such that each tracked-edit in the reconstruction process can progressively be tested. Users edit the model with their prefered reconstruction tool, and export to SBML afterwards. This way each incremental change can be tested with the suite. Then, a report on the entire history of results serves as a guide towards a functional, high quality GEM. This workflow is accessible through the command line, and may be extended to include custom tests against experimental data. Memote allows researchers to test a model repository offline, but we encourage supporting community collaboration in reconstruction via GitHub or other web-based version control repositories such as GitLab (https://gitlab.com/).

Using a branching strategy (Figure 4-1b), model builders could curate different parts of the model simultaneously, invite external experts to comment on model features via GitHub's issue system, or allow external partners to make pull-requests with improvements to individual pathways or supplying data. This further allows model authors to act as gatekeepers, choosing to accept only high quality contributions. Identification of functional differences happens in the form of a comparative report, while for the file-based differences memote capitalizes on GitHub's ability to show the line-by-line changes between different versions of a model's SBML file.

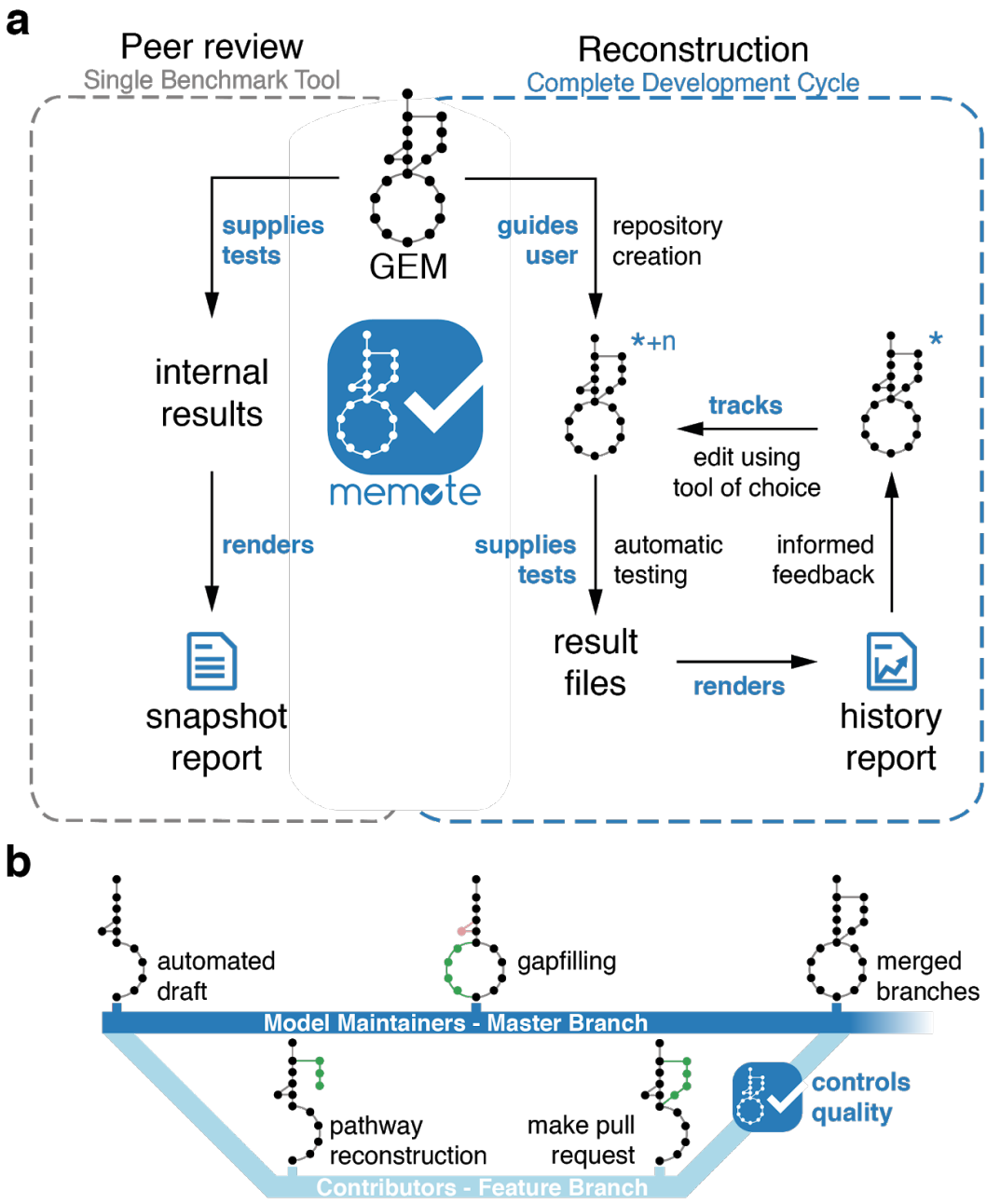


Figure 4-1. Functionality offered by memote. (a) Graphical representation of the two principal workflows in detail. For peer review, memote serves as a benchmark tool offering a quick snapshot report. For model reconstruction, memote guides the user in creating a version-controlled repository for the model (indicated by the *****), and in activating continuous integration. The model is tested using memote's library of tests, the results are saved and an initial report of the model is generated. This constitutes the first iteration of the development cycle. Now, the user may edit the model using a tool of their choice creating a new version (indicated by the **+n**). This will restart the cycle by running the tests automatically, saving the results for each version and including them incrementally in a history report. (b) An example of a potential branching strategy employing memote as a benchmark of external contributions. **Bold blue text** denotes actions performed by memote.

4.3.3. Description of the test library

The tests within memote are divided into independent core tests and tests that depend on user-supplied experimental data. Core tests are further divided into a scored and an unscored section (Figure 4-2).

The tests in the scored section are independent of the type of organism that is being modeled, the complexity of the model itself or the types of identifiers that are used to describe the model components. Calculating a score for these tests allows for the quick comparison of any two given models at a glance. The unscored section provides general statistics and covers specific aspects of a model that are not universally applicable. For instance, dedicated quality control of the biomass equation only applies to metabolic models which are used to investigate cell growth. Test in either section generally belong to one of four areas:

1) Basic tests give an insight into the formal correctness of a model, verifying the existence of the main model components such as metabolites, compartments, reactions, and genes. These tests also check for the presence of formula and charge information of metabolites, and for the presence of gene-protein-reaction rules of reactions. General quality metrics such as the degree of metabolic coverage representing the ratio of reactions and genes (Monk, Nogales, & Palsson, 2014) are also covered here.

2) The biomass reaction is based on the biomass composition of the modeled organism and expresses its ability to produce the necessary precursors for *in silico* cell growth and maintenance. Hence, an extensive, well-formed biomass reaction is crucial for accurate predictions with a GEM (Xavier et al., 2017). Thus, a number of tests are dedicated to testing the biomass reaction. This includes testing the model's ability to produce all biomass precursors in different conditions, the biomass consistency, a non-zero growth rate and direct precursors.

3) Unbalanced reactions, stoichiometric inconsistency, erroneously produced energy metabolites (Fritzemeier et al. 2017) and permanently blocked reactions, are identified by testing the model's consistency. Errors here may lead to the production of ATP or redox cofactors from nothing (Thiele & Palsson, 2010) and are detrimental to the performance of the model when using FBA (Ravikrishnan & Raman, 2015).

4) Annotation tests maintain that a model is equipped according to the community standards with MIRIAM-compliant cross-references (Le Novère et al., 2005), that all primary IDs belong to the same namespace as opposed to being fractured across several namespaces, and that components are described semantically with Systems Biology Ontology terms (Courtot et al. 2011). A lack of explicit, standardized annotations complicates the use, comparison and extension of GEMs, and thus strongly hampers collaboration (Dräger & Palsson, 2014; Heavner & Price, 2015; Ravikrishnan & Raman, 2015).

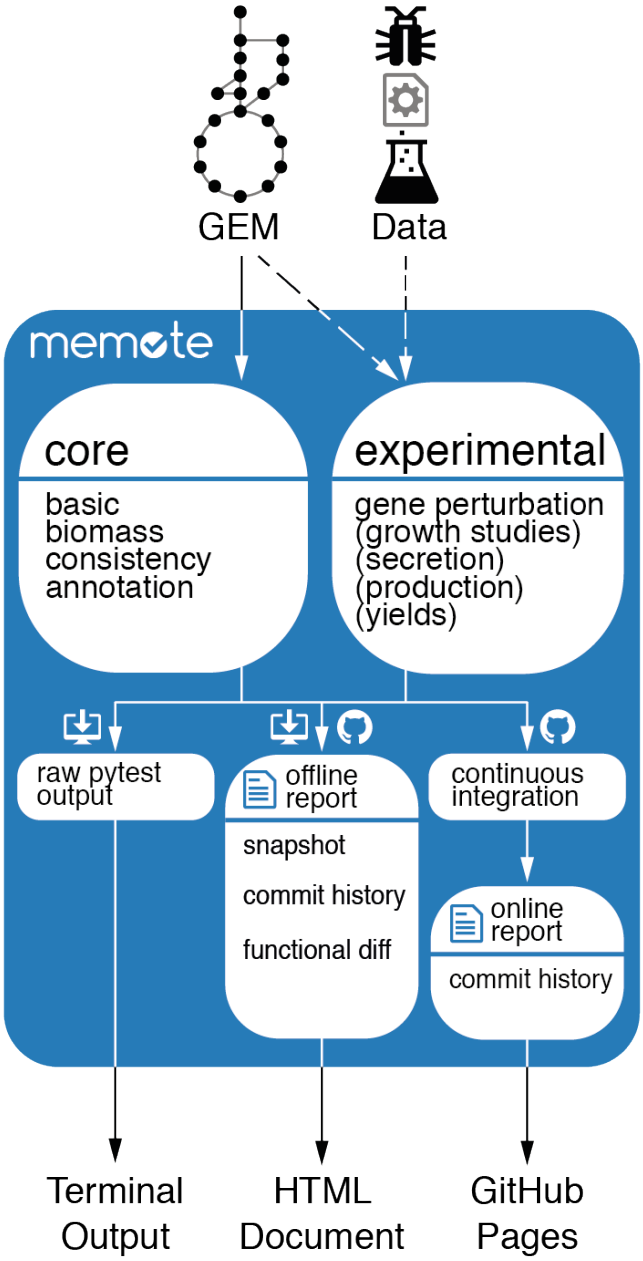


Figure 4-2. Functional overview of the Metabolic Model Tests (memote) package: A genome-scale metabolic model (GEM) is supplied by the user and tested in all core test categories. Optionally, if the user supplies experimental data, the model will also be subjected to the corresponding experimental tests. After testing, user input through the command line interface determines how the results are displayed. In addition to a high-level output in the terminal, the user can generate a variety of HTML-formatted reports. "Snapshot" will provide a performance benchmark of a single specified model; with a functional diff the user can benchmark two models side-by-side; and the commit-history will show the development of a model's performance over the course of changes to a version controlled model. The latter is the type of report that is generated automatically when continuous integration is enabled. Then the results are displayed online on the project's GitHub pages. Future features are denoted as **(feature)**.

In addition to the core tests, researchers may supply experimental data from gene perturbation studies from a range of input formats (CSV, TSV, XLS, XLSX, or ODS). Gene perturbation studies, especially gene essentiality studies are useful to refine GEM reconstructions by allowing researchers to identify network gaps and by providing a basis for model validation (Feist et al. 2009), as well as providing grounds for hypothesis about an organism's physiology (Oberhardt, Palsson, & Papin, 2009).

In order to constrain the model with respect to the experimental conditions underlying the supplied data, researchers may optionally define a configuration file (.yml) in which they can set the medium, FBA objective, and known regulatory effects. Without memote, this would typically be done through the use of custom scripts, which can vary significantly depending on the researcher writing them. Moreover, scripts tend to suffer from software rot if they are not actively maintained after publication (Beaulieu-Jones and Greene 2017). The use of configuration files instead of scripts avoids software rot, since the configuration files do not require other dependencies than memote, which is likely to be maintained in the future. In conjunction, setting up a version-controlled model repository, not only allows researchers to publish a 'default' unspecific GEM of the investigated organism, but also reproducible instructions on how to obtain a model that is specific to the organism in a defined experimental context including, and validated against the data supporting this context. This formulaic approach of deriving a GEM into a condition-specific form supports Heavner and Price's (2015) call for more transparency and reproducibility in metabolic network reconstruction.

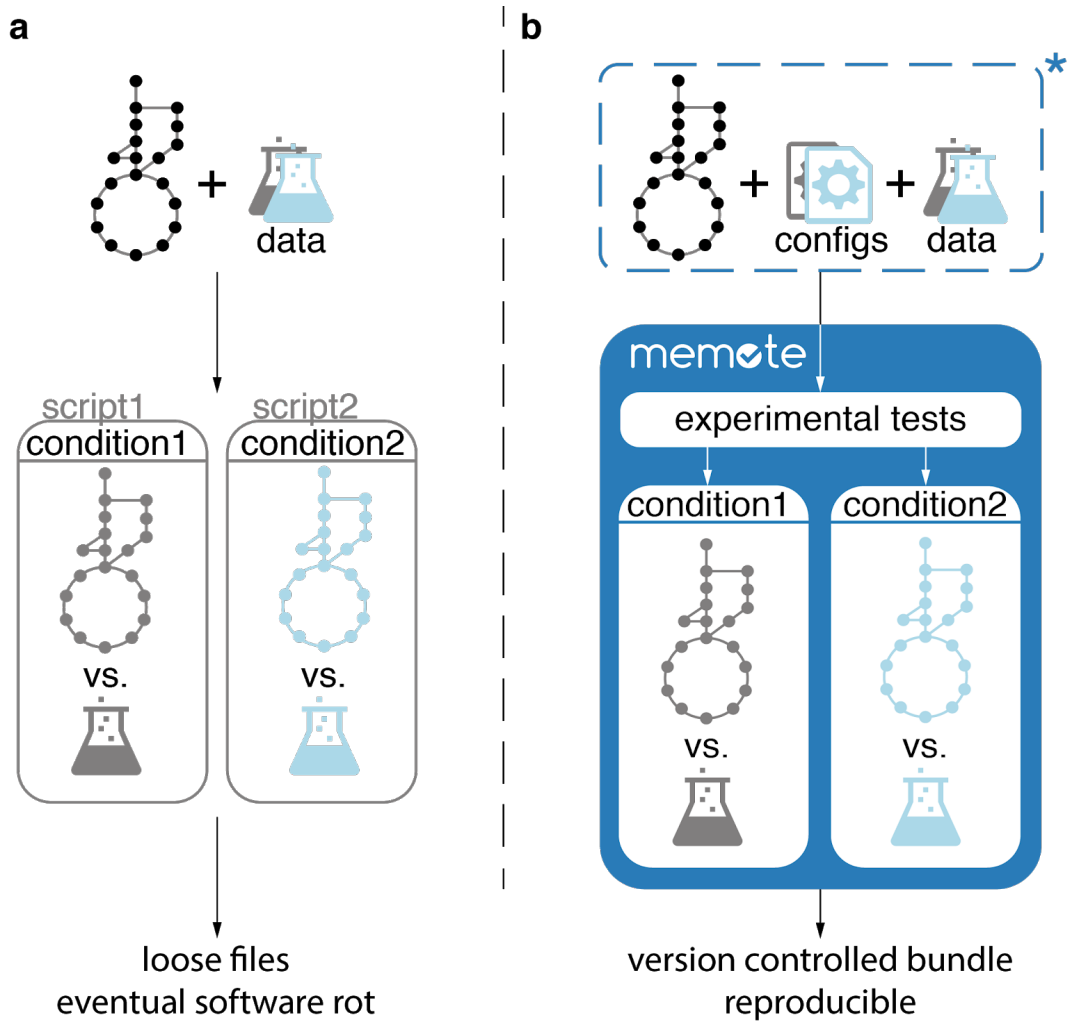


Figure 4-3. Experimental tests can be tailored to a specific condition through the use of one or several configuration files (config). (a) To validate GEMs against experimental data measured in specific conditions, researchers usually write their own scripts which constrain the model. This is problematic as scripts can vary a lot and they are, unless actively maintained, susceptible to software rot. (b) With memote, pre-defined configuration files replace scripts, which allows the experimental validation of GEMs to be unified and formalized. Bundling the model, configuration files and experimental data within a version-controlled repository (indicated by the *) supports cohesive releases.

4.4. Discussion

By providing a performance benchmark based on community guidelines and commonly-referenced SOPs, memote facilitates informed model reconstruction

and quality control. The tests within memote cover semantic and conceptual requirements which are fundamental to SBML3FBC and constraints-based modeling, respectively. They are extensible to allow the validation of a model's performance against experimental data, and can be executed as a stand-alone tool or integrated into existing reconstruction pipelines. Capitalizing on robust workflows established in modern software development, memote promotes openness and collaboration by granting the community tangible metrics to support their research and to discuss assumptions or limitations openly.

The concept of having a set of defined metabolic model tests is not dependent on the implementation in memote presented herein. In fact, for some platforms it may be more desirable to implement these tests separately as this could streamline the user experience. However, an independent, central, community-maintained library of tests and a tool to run them offers 1) an unbiased approach to quality control as the tests are continuously reviewed by a large group of people, 2) a long-lived resource as the project is independent of individual funding sources, 3) flexibility as updates can be propagated rapidly and 4) consistent results as the codebase is unified. To encourage integration as opposed to duplication, memote provides a python API as well as being available as a web-service. We plan to make memote available in the Department of Energy's Knowledge Base (Arkin et al. 2016) as an app and integrate it with the BiGG Database (King et al. 2016), BioModels (Chelliah et al. 2015) and the RAVEN toolbox (Agren et al. 2013).

The variety of constraints-based modeling approaches and the fundamental differences between various organisms compound the assessment of GEMs. For instance, authors may publish metabolic networks, which are constrained to reflect one experimental condition, or publish unconstrained metabolic databases, which need to be initialized before simulation. Both can be encoded in SBML. With having a scored test section, we attempt to normalize each type of model such that they become comparable. Despite memote's code itself being unit tested, it is difficult to anticipate all edge cases *a priori*. In addition, memote

depends on external resources such as MetaNetX (Moretti et al. 2016) and identifiers.org (Juty et al. 2012) that are likely to change over time. Subsequently, individual users may identify potential false-positive and false-negative results. Hence, we recommend approaching the report with scrutiny and encourage users to reach out to the authors of the manuscript to report any errors.

The tests in memote only apply to SBML-encoded, stoichiometric models. However, the underlying principles and individual tests behind memote may be applicable to models of metabolism and expression (ME-models) (O'Brien, Lerman, Chang, Hyduke, & Palsson, 2013), kinetic (Vasilakou et al. 2016), or even systems pharmacological models (Thiel et al. 2017). In the context of rising big data streams, memote ought to be extended to provide support for tests based on multi-omics data. Furthermore, to distribute all files of a model repository together i.e. the model, supporting data and scripts, these could be automatically bundled into one COMBINE archive file (Bergmann et al. 2014).

The greater flexibility and awareness of community-driven, open-source development and the trend towards modular approaches exhibited by the solutions that were put forth in the field of systems biology (Dräger & Palsson, 2014), motivate us to keep the development of memote open. We believe that a robust benchmark can only come to fruition when actively supported by the whole community and thus call for interested experts to involve themselves, be it through testing our tool, discussing its content or improving its implementation. We intend to keep extending memote with additional tests and functionality.

4.5. Acknowledgements

The authors acknowledge financial support from Innovation Fund Denmark (project "Environmentally Friendly Protein Production (EFPro2)") and the Novo Nordisk Foundation.

NS, MB, and MH: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 686070 (DD-DeCaF).

4.6. References

Baker, M. (2016). How quality control could save your science. *Nature*, *529*(7587), 456–458. https://doi.org/10.1038/529456a

Bergmann, F. T., & Sauro, H. M. (2006). SBW - A Modular Framework for Systems Biology. In *Proceedings of the 2006 Winter Simulation Conference* (pp. 1637–1645). https://doi.org/10.1109/WSC.2006.322938

Bordbar, A., Feist, A. M., Usaite-Black, R., Woodcock, J., Palsson, B. O., & Famili, I. (2011). A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Systems Biology*, *5*, 180. https://doi.org/10.1186/1752-0509-5-180

Bordbar, A., Monk, J. M., King, Z. A., & Palsson, B. O. (2014). Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews. Genetics*, *15*(2), 107–120. https://doi.org/10.1038/nrg3643

Bornstein, B. J., Keating, S. M., Jouraku, A., & Hucka, M. (2008). LibSBML: an API library for SBML. *Bioinformatics* , *24*(6), 880–881. https://doi.org/10.1093/bioinformatics/btn051

Brindescu, C., Codoban, M., Shmarkatiuk, S., & Dig, D. (2014). How Do Centralized and Distributed Version Control Systems Impact Software Changes? In *Proceedings of the 36th International Conference on Software Engineering* (pp. 322–333). New York, NY, USA: ACM. https://doi.org/10.1145/2568225.2568322

Chan, S. H. J., Cai, J., Wang, L., Simons-Senftle, M. N., & Maranas, C. D. (2017). Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics* . https://doi.org/10.1093/bioinformatics/btx453

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K., & Reuss, M. (2002). Dynamic modeling of the central carbon metabolism of Escherichia coli. *Biotechnology and Bioengineering*, *79*(1), 53–73. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17590932

Chelliah, V., Juty, N., Ajmera, I., Ali, R., Dumousseau, M., Glont, M., … Laibe, C. (2015). BioModels: ten-year anniversary. *Nucleic Acids Research*, *43*(Database issue), D542–8. https://doi.org/10.1093/nar/gku1181

Dick, T., Manjunatha, U., Kappes, B., & Gengenbacher, M. (2010). Vitamin B6 biosynthesis is essential for survival and virulence of Mycobacterium tuberculosis. *Molecular Microbiology*, *78*(4), 980–988. https://doi.org/10.1111/j.1365-2958.2010.07381.x

Dräger, A., & Palsson, B. Ø. (2014). Improving collaboration by standardization efforts in systems biology. *Frontiers in Bioengineering and Biotechnology*, *2*, 61. https://doi.org/10.3389/fbioe.2014.00061

Dräger, A., Rodriguez, N., Dumousseau, M., Dörr, A., Wrzodek, C., Le Novère, N., … Hucka, M. (2011). JSBML: a flexible Java library for working with SBML. *Bioinformatics* , *27*(15), 2167–2168. https://doi.org/10.1093/bioinformatics/btr361

Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., … Palsson, B. Ø. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(6), 1777–1782. https://doi.org/10.1073/pnas.0610772104

Duarte, N. C., Herrgård, M. J., & Palsson, B. Ø. (2004). Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*, *14*(7), 1298–1309. https://doi.org/10.1101/gr.2250904

Durot, M., Bourguignon, P. Y., & Schachter, V. (2009). Genome-scale models of bacterial metabolism: Reconstruction and applications. *FEMS Microbiology Reviews*, *33*(1), 164–190. https://doi.org/10.1111/j.1574-6976.2008.00146.x

Duvall, P. M., Matyas, S., & Glover, A. (2007). *Continuous Integration: Improving Software Quality and Reducing Risk*. Pearson Education. Retrieved from https://market.android.com/details?id=book-PV9qfEdv9L0C

Ebrahim, A., Almaas, E., Bauer, E., Bordbar, A., Burgard, A. P., Chang, R. L., … Thiele, I. (2015). Do genome-scale models need exact solvers or clearer standards? *Molecular Systems Biology*, *11*(10), 831. https://doi.org/10.15252/msb.20156157

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, *7*, 74. https://doi.org/10.1186/1752-0509-7-74

Edwards, J. S., & Palsson, B. O. (1999). Systems properties of the Haemophilus influenzae Rd metabolic genotype. *The Journal of Biological Chemistry*, *274*(25), 17410–17416. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10364169

Edwards, J. S., & Palsson, B. O. (2000). The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(10), 5528–5533. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/10805808

Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L., & Palsson, B. O. (2009). Reconstruction of Biochemical Networks in Microbial Organisms. *Nature Reviews. Microbiology*, *7*(2), 129–143. https://doi.org/10.1038/nrmicro1949.Reconstruction

Förster, J., Famili, I., Palsson, B. O., & Nielsen, J. (2003). Large-scale evaluation of in silico gene deletions in Saccharomyces cerevisiae. *Omics: A Journal of Integrative Biology*, *7*(2), 193–202. https://doi.org/10.1089/153623103322246584

Fowler, M., & Foemmel, M. (2006). Continuous integration. *Thought-Works) Http://www. Thoughtworks. com/Continuous Integration. Pdf*, *122*. Retrieved from http://www.dccia.ua.es/dccia/inf/asignaturas/MADS/2013-14/lecturas/10_Fowler_Continuous_Integration.pdf

Frank T. Bergmann, S. M. K. (2011, September 12). *SBML Team facilities & software*. https://doi.org/10.1038/npre.2011.6401.1

Gevorgyan, A., Poolman, M. G., & Fell, D. A. (2008). Detection of stoichiometric inconsistencies in biomolecular models, *24*(19), 2245–2251. https://doi.org/10.1093/bioinformatics/btn425

Heavner, B. D., & Price, N. D. (2015). Transparency in metabolic network reconstruction enables scalable biological discovery. *Current Opinion in Biotechnology*, *34*, 105–109. https://doi.org/10.1016/j.copbio.2014.12.010

Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., … Lewis, N. E. (2016). A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Systems*, *3*(5), 434–443.e8. https://doi.org/10.1016/j.cels.2016.10.020

Hucka, M., Bergmann, F. T., Hoops, S., Keating, S. M., Sahle, S., Schaff, J. C., … Wilkinson, D. J. (2010). The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core, (L3 V1 Core), 167. Retrieved from http://authors.library.caltech.edu/50974/

Jensen, K., Cardoso, J., & Sonnenschein, N. (2016). Optlang: An algebraic modeling language for mathematical optimization. *Journal of Open Source Software*. Retrieved from http://www.forskningsdatabasen.dk/en/catalog/2350686580

Jerby, L., Shlomi, T., & Ruppin, E. (2010). Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Molecular Systems Biology*, *6*, 401. https://doi.org/10.1038/msb.2010.56

Kell, D. B., & Goodacre, R. (2014). Metabolomics and systems pharmacology: why and how to model the human metabolic network for drug discovery. *Drug Discovery Today*, *19*(2), 171–182. https://doi.org/10.1016/j.drudis.2013.07.014

Khandelwal, R. A., Olivier, B. G., Röling, W. F. M., Teusink, B., & Bruggeman, F. J. (2013). Community flux balance analysis for microbial consortia at balanced growth. *PloS One*, *8*(5), e64567. https://doi.org/10.1371/journal.pone.0064567

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., … Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, *44*(D1), D515–22. https://doi.org/10.1093/nar/gkv1049

Le Novère, N., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., … Wanner, B. L. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature Biotechnology*, *23*(12), 1509–1515. https://doi.org/10.1038/nbt1156

Lewis, N. E., Schramm, G., Bordbar, A., Schellenberger, J., Andersen, M. P., Cheng, J. K., … Palsson, B. Ø. (2010). Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nature Biotechnology*, *28*(12), 1279–1285. https://doi.org/10.1038/nbt.1711

Mahadevan, R., & Henson, M. A. (2012). Genome-based Modeling and Design of Metabolic Interactions in Microbial Communities. *Computational and Structural Biotechnology Journal*, *3*, e201210008. https://doi.org/10.5936/csbj.201210008

Meyer, M. (2014). Continuous Integration and Its Tools. *IEEE Software*, *31*(3), 14–16. https://doi.org/10.1109/MS.2014.58

Monk, J., Nogales, J., & Palsson, B. O. (2014). Optimizing genome-scale network reconstructions. *Nature Biotechnology*, *32*(5), 447–452. https://doi.org/10.1038/nbt.2870

Nagappan, N., Michael Maximilien, E., Bhat, T., & Williams, L. (2008). Realizing quality improvement through test driven development: results and experiences of

four industrial teams. *Empirical Software Engineer*, *13*(3), 289–302. https://doi.org/10.1007/s10664-008-9062-z

Oberhardt, M. A., Palsson, B. Ø., & Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*, *5*(1), 320. https://doi.org/10.1038/msb.2009.77

O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., & Palsson, B. Ø. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, *9*, 693. https://doi.org/10.1038/msb.2013.52

Oh, Y. K., Palsson, B. O., Park, S. M., Schilling, C. H., & Mahadevan, R. (2007). Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. *The Journal of Biological Chemistry*, *282*(39), 28791–28799. https://doi.org/10.1074/jbc.M703759200

Olivier, B. G., & Bergmann, F. T. (2015). The Systems Biology Markup Language (SBML) Level 3 Package: Flux Balance Constraints. *Journal of Integrative Bioinformatics*, *12*(2), 269. https://doi.org/10.2390/biecoll-jib-2015-269

Olivier, B. G., Swat, M. J., & Moné, M. J. (2016). Modeling and Simulation Tools: From Systems Biology to Systems Medicine. *Methods in Molecular Biology* , *1386*, 441–463. https://doi.org/10.1007/978-1-4939-3283-2_19

Osborne, J. M., Bernabeu, M. O., Bruna, M., Calderhead, B., Cooper, J., Dalchau, N., … Deane, C. (2014). Ten simple rules for effective computational research. *PLoS Computational Biology*, *10*(3), e1003506. https://doi.org/10.1371/journal.pcbi.1003506

Ravikrishnan, A., & Raman, K. (2015). Critical assessment of genome-scale metabolic networks: the need for a unified standard. *Briefings in Bioinformatics*, *16*(6), 1057–1068. https://doi.org/10.1093/bib/bbv003

Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., … Palsson, B. Ø. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, *6*(9), 1290–1307. https://doi.org/10.1038/nprot.2011.308

Schilling, C. H., Covert, M. W., Famili, I., Church, G. M., Edwards, J. S., & Palsson, B. O. (2002). Genome-scale metabolic model of Helicobacter pylori 26695. *Journal of Bacteriology*, *184*(16), 4582–4593. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12142428

Sheikh, K., Förster, J., & Nielsen, L. K. (2005). Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus. *Biotechnology Progress*, *21*(1), 112–121. Retrieved from http://onlinelibrary.wiley.com/doi/10.1021/bp0498138/full

Thiele, I., & Palsson, B. Ø. (2010, January). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*. Nature Publishing Group. https://doi.org/10.1038/nprot.2009.203

Vasilakou, E., Machado, D., Theorell, A., Rocha, I., Nöh, K., Oldiges, M., & Wahl, S. A. (2016). Current state and challenges for dynamic metabolic modeling. *Current Opinion in Microbiology*, *33*, 97–104. https://doi.org/10.1016/j.mib.2016.07.008

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Xavier, J. C., Patil, K. R., & Rocha, I. (2017). Integration of Biomass Formulations of Genome-Scale Metabolic Models with Experimental Data Reveals Universally Essential Cofactors in Prokaryotes. *Metabolic Engineering*, *39*, 200–208. https://doi.org/10.1016/j.ymben.2016.12.002

Yuan, Q., Huang, T., Li, P., Hao, T., Li, F., Ma, H., … Goryanin, I. (2017). Pathway-Consensus Approach to Metabolic Network Reconstruction for Pseudomonas putida KT2440 by Systematic Comparison of Published Models. *PloS One*, *12*(1), e0169437. https://doi.org/10.1371/journal.pone.0169437

Zomorrodi, A. R., & Maranas, C. D. (2012). OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Computational Biology*, *8*(2), e1002363. https://doi.org/10.1371/journal.pcbi.1002363

Chapter 5. Modeling Methanotrophy: A genome-scale reconstruction of *Methylococcus capsulatus* (Manuscript 3)

This chapter represents a manuscript that is currently in preparation. It thus deviates from the introductory chapters in that its structure follows the typical format of a scientific paper. It has been authored by Christian Lieven, Krist V. Gernaey, Markus J. Herrgard, and Nikolaus Sonnenschein. Christian Lieven conducted most of the work.

5.1. Abstract

Genome-scale metabolic models allow researchers to investigate the metabolism of a given organism in various growth conditions. In addition, they provide a means to calculate yields, to predict consumption and production rates, and to study the effect of genetic modifications, without running resource-intensive experiments. While metabolic models have become an invaluable tool for optimizing industrial production hosts like *E. coli* and *S. cerevisiae*, few such models exist for C1 metabolizers. Here we present a genome-scale metabolic model for *Methylococcus capsulatus*, a well-studied obligate methanotroph, which, since the 70s has been the industry's focus as a production strain of single cell protein (SCP). The model was manually curated, and spans a total of 877 metabolites connected via 898 reactions. The inclusion of 730 genes and a host of annotations, make this model not only a useful tool for knockout studies, but also a centralized knowledge base for M. capsulatus. We are confident that our contribution will serve the ongoing fundamental research of C1 metabolism, and pave the way for rational strain design strategies towards improved SCP production in *M. capsulatus*.

5.2. Introduction

The Gram-negative, obligate-aerobe *Methylococcus capsulatus* is a methane oxidizing, gamma-proteobacterium. Since its initial isolation by Foster and Davis (1966), the organism has been subject to a wide array of studies. The global role of *M. capsulatus* as a participant in the carbon cycle has been elucidated (Hanson & Hanson, 1996; Jiang et al., 2010) as well as its effects on human (Indrelid, Kleiveland, Holst, Jacobsen, & Lea, 2017) and animal health and disease (Kleiveland et al., 2013). Specifically the latter studies have been triggered by a considerable commercial interest in *M. capsulatus* as the primary microbe used for the production of Single Cell Protein (SCP) as animal feed starting in the 70s (Margareth Øverland, Tauson, Shearer, & Skrede, 2010). Now that natural gas is recognised as a cheap and abundant alternative (Ritala, Häkkinen, Toivari, & Wiebe, 2017), the applications of *M. capsulatus* as a key part in this technology are being explored again (Nunes et al., 2016; Petersen et al., 2017). The largest portion of studies however have focused on uncovering the organism's biochemistry: Its unique metabolism which combines both the ribulose monophosphate pathway (RuMP) and a partial serine pathway for formate assimilation have garnered some attention, but the greatest interest has been the role and function of the initial enzyme in methanotrophy, the methane monooxygenase (Ross & Rosenzweig, 2017).

Methylococcus capsulatus is able to express two distinct types of methane monooxygenases: a soluble form of methane monooxygenase (sMMO) and a particulate, or membrane-bound form (pMMO). The expression of these enzymes is strongly influenced by the concentration of copper in the medium; when *M. capsulatus* is grown in low levels of copper the sMMO is active, while the pMMO is predominantly active in high levels. Both enzymes require an external electron donor to break a C-H bond in methane. The genome sequence published by (Ward et al., 2004) revealed the existence of several cytochrome proteins which may contribute to a dedicated electron transfer chain which may contribute to methane oxidation (Trotsenko & Murrell, 2008). This is supported by

transcriptomic profiling which attributes a majority of genes that are differentially expressed in varying copper concentrations to energy and transport processes (Larsen & Karlsen, 2016). While the electron donor for the sMMO is NADH (Blazyk & Lippard, 2002; Lund, Woodland, & Dalton, 1985), the native reductant to the pMMO has not yet been elucidated due to difficulties to purify and assay samples of the enzyme (Ross & Rosenzweig, 2017).

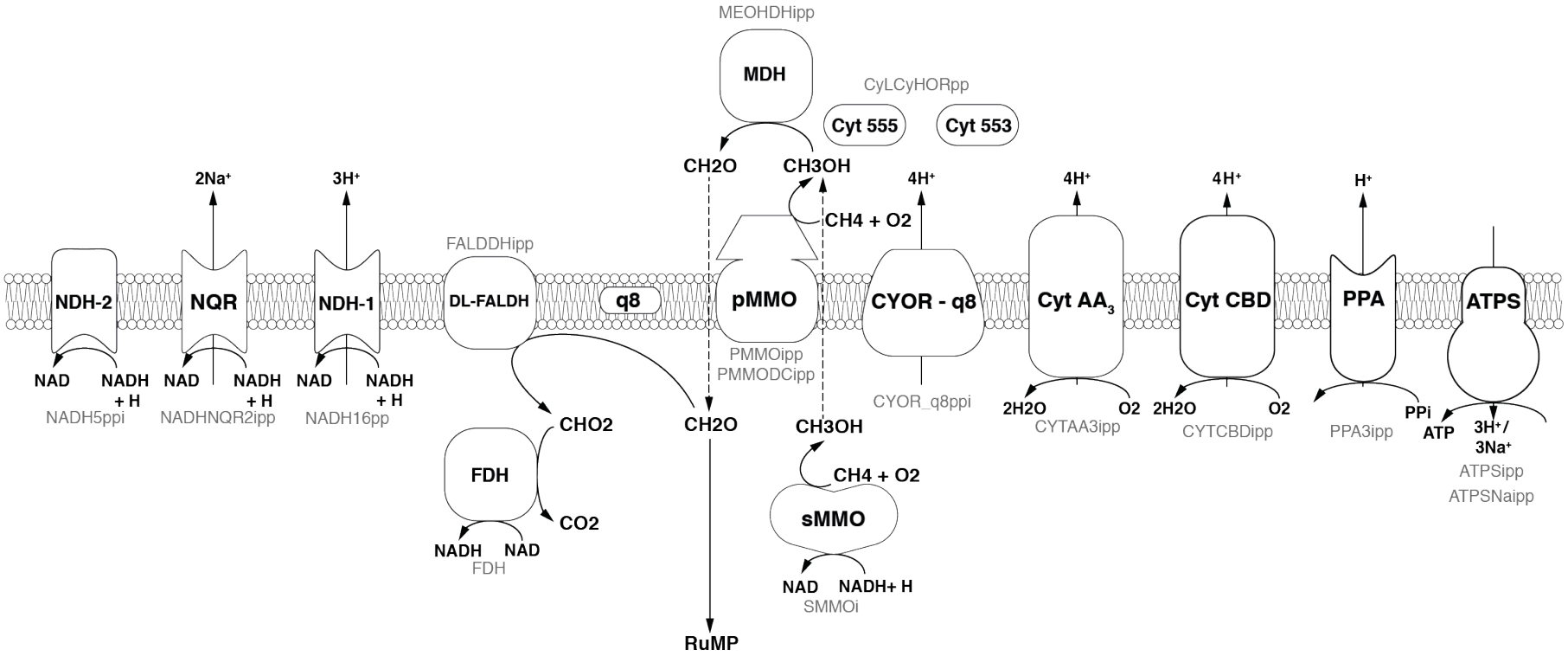


Figure 5-1. Overview of the respiratory chain in *Methylococcus capsulatus* as implemented in iCL730. Black text in the center of the symbols denotes the common abbreviation, while the faint grey text below denotes the corresponding reaction ID in the metabolic model.

Three possible modes of electron transfer to the pMMO have been proposed.

1) In the *redox-arm* mode (Dawson and Jones 1981), the methanol dehydrogenase (MDH) passes electrons via cytochrome c555 (cL) (Christopher Anthony, 1992) and cytochrome c553 (cH) (DiSpirito, Kunz, Choi, & Zahn, 2004) to either a CBD- or AA3-type terminal oxidase (Larsen & Karlsen, 2016) and thus contribute to building up a proton motive force (PMF) and the synthesis of ATP (Figure 1). The electrons required for the oxidation of methane are provided through ubiquinone (Q8H2), which in turn recieved electrons from various dehydrogenases involved in the oxidation of formaldehyde to CO_2. Although no binding site has been identified, there is support for pMMO reduction by endogenous quinols (Shiemke, Arp, & Sayavedra-Soto, 2004).

2) With the *direct coupling* mode, the MDH is able to directly pass electrons to the pMMO (Culpepper and Rosenzweig 2014; Leak and Dalton 1983; Wolfe and Higgins 1979). Here, cytochrome c555 is the electron donor to the pMMO instead of ubiquinol. This mode is supported by results from cryoelectron microscopy which indicates that the pMMO and the MDH form a structural enzyme complex (Myronova et al. 2006).

3) The *uphill electron transfer* mode supposes that the electrons from cytochrome c553 can reach the ubiquinol-pool facilitated by the PMF at the level of the ubiquinol-cytochrome-c reductase (CYOR_q8). This mode was proposed by Leak & Dalton (1986b) as it could explain the observed reduced efficiency.

To determine which of these modes is most likely active in *M. capsulatus,* Leak and Dalton (1986) developed mathematical models for each mode based on previous calculations by Anthony (1978), and Harder and Van Dijken (1976). However, their models did not reflect the experimental results. De la Torre et al. (2015) constructed a genome-scale metabolic model (GEM) for *Methylomicrobium buryatense* to investigate growth yields and energy requirements in different conditions. They found that the *redox-arm* mode correlated least with their experimental data for *M. buryatense*.

A GEM not only represents a knowledge base that combines the available physiological, genetic and biochemical information of a single organism (Thiele and Palsson 2010), it also provides a testbed for rapid prototyping of a given hypothesis (Benedict et al. 2012). Here, we present the first manually curated genome-scale metabolic model for *Methylococcus capsulatus,* with the intent of supplying the basis for hypothesis-driven metabolic discovery and clearing the way for future efforts in metabolic engineering (Kim et al. 2015). We validated the model by fitting it to experimental data from Leak and Dalton (1986), compare it to the model of *M. buryatense* (de la Torre et al. 2015) and explain notable differences by considering the the proposed electrons transfer modes.

5.3. Results and Discussion

5.3.1. Reconstruction

The presented genome-scale metabolic reconstruction of *M. capsulatus* termed iCL730 was based on BMID000000141026, an automatic draft published as part of the Path2Models project (Büchel et al., 2013). The whole genome sequence of *Methylococcus capsulatus* (Bath) (GenBank AE017282.2 (Ward et al., 2004)) was used to aid the curation process and to supply annotations (see material & methods). This metabolic reconstruction consists of 840 enzymatic reactions that interconvert 783 unique metabolites. The total number of reactions, including exchange, demand and the biomass reactions, amounts to 892. The model attributes reactions and metabolites to three distinct compartments: The cytosol, the periplasm, and the medium that is referred to as extracellular in the model. 730 unique ORFs support 85.5% of the included reactions with Gene-Protein-Reaction rules (GPR), leaving 122 reactions without GPR. On the basis of GPR, we were able to identify 24 enzyme complexes.

The model is made available in the community-standard SMBL format (Level 3 Version 2 with FBC).

5.3.2. Biomass reaction

In stoichiometric models biological growth is represented as the flux through a special demand reaction. The so-called biomass reaction functions as a drain of metabolites, which are either highly reduced non-polymeric macromolecules such as lipids and steroids or precursors to typical biopolymers such as nucleic acids, proteins or long-chain carbohydrates. The stoichiometry of an individual precursor was calculated from the principal composition of *M. capsulatus as* provided by Unibio A/S (at www.unibio.dk). The monomer composition of individual macromolecules was calculated from different sources. A detailed account of the resources is provided in the methods section and an overview of the biomass reaction is given in Supplemental Table 1.

As experimental data was not available, the growth-associated maintenance requirement (GAM) amounting to 23.087 mmol ATP gDW-1 h-1 was estimated according to Thiele et al. (2010) based on the data for *E. coli* published by Neidhardt et al. (1990). The value for GAM is expected to increase with the growth rate of the cells (Varma et al. 1993). Like de la Torre et al. (2015) had done for *M. buryatense*, we assumed the non-growth associated maintenance (NGAM) of *M. capsulatus* to be similar to that of *E. coli* thus setting it to 8.39 mmol ATP gDW-1 h-1 (Feist et al., 2010). In either case, that of GAM and that of NGAM, the specific requirements for *M. capsulatus* are yet to be determined experimentally.

5.3.3. Metabolism

Much of the focus in curating the initial draft model BMID000000141026 was put on the central carbon metabolism and respiration of *M. capsulatus*.

Energy-dependent oxidation of methane is the first step in aerobic methanotrophy. While it has been established that the sMMO receives electrons from NADH (Colby & Dalton, 1978, 1979) generated along the oxidation of formaldehyde to CO_2, it is still not entirely clear whether the pMMO receives

electrons exclusively from the ubiquinol-pool (Choi et al., 2003; DiSpirito et al., 2004). To study which of the three modes of electron transfer is active in *M. capsulatus*, they were implemented as follows (Figure 2). To include the *redox-arm* we implemented the reaction representing the particulate methane monooxygenase, in the model termed as PMMOipp, utilizing Q8H2 as a cofactor. Accordingly, a variant of the pMMO reaction was added to account for a possible direct coupling to the MDH. In this variant reaction, termed PMMODCipp, the cofactor is cytochrome c555, represented as the metabolite focytcc555_p in the model. To enable an *uphill electron transfer*, the reaction representing the ubiquinol-cytochrome-c reductase (CYOR_q8ppi in the model), was constrained to be reversible while keeping PMMOipp active.

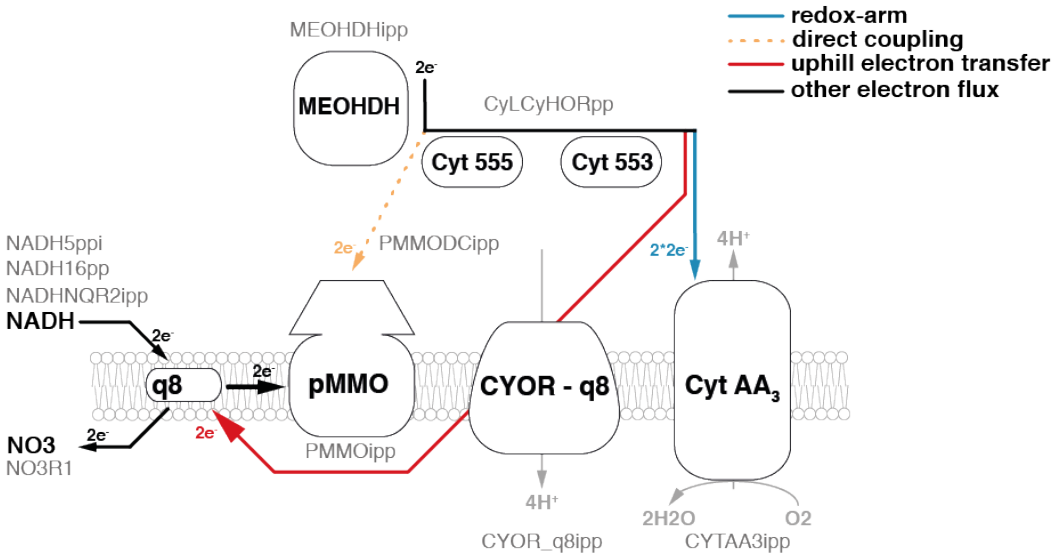


Figure 5-2. The three possible modes of electron transfer to the pMMO mapped onto an excerpt of the respiratory chain. 1) Redox-arm: The methanol dehydrogenase transfers electrons to the terminal oxidase, while the pMMO draws electrons from the quinone pool. 2) Direct coupling: Electrons from the oxidation of methanol are transferred directly to the pMMO. 3) Uphill electron transfer: Electrons from the methanol dehydrogenase feed back into the ubiquinol-pool. Black text denotes the common name, while faint grey text denotes the reaction ID in the metabolic model.

Following the path of carbon through metabolism downstream from the MDH, the model includes both the reaction for a ubiquinone-dependent formaldehyde

dehydrogenase (Zahn, Bergmann, Boyd, Kunz, & DiSpirito, 2001), termed FALDDHipp, and an NAD-dependent version, termed ALDD1 (Figure 3). Despite of the initial evidence for the latter reaction (Tate & Dalton, 1999) having been dispelled by Adeosun (2004), it was added to allow further investigation into the presence of a putative enzyme of that function. An additional pathway for formaldehyde oxidation represented in both genome and model is the tetrahydromethanopterin(THMPT)-linked pathway (Vorholt, 2002).

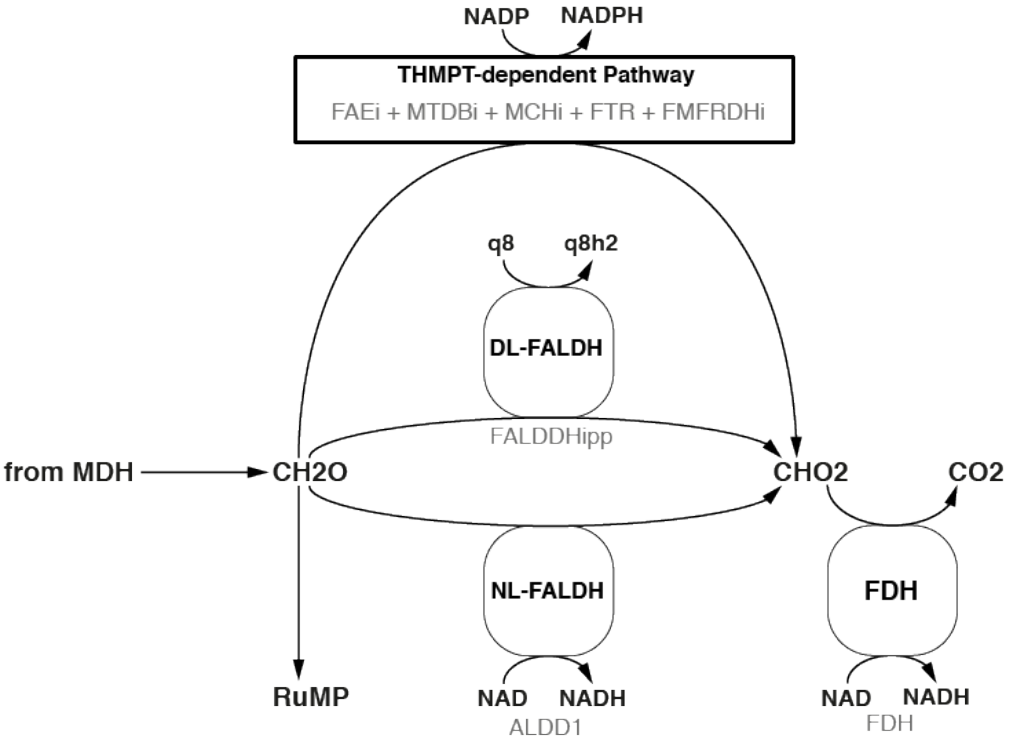


Figure 5-3. Three different formaldehyde oxidation pathways are represented in the model. Black text denotes the common name, while faint grey text denotes the reaction ID in the metabolic model.

Formaldehyde assimilation in *M. capsulatus* occurs primarily through the ribulose monophosphate (RuMP)-pathway. As outlined by Anthony (1983), the RuMP-

pathway has four hypothetical variants. Based on the annotated genome published by Ward et al. (2004), we identified not only both C6 cleavage pathways depending either on the 2-keto, 3-deoxy, 6-phosphogluconate (KDPG) aldolase (EDA) or the fructose bisphosphate aldolase (FBA), but also the transaldolase (TA) involved in the rearrangement phase that regenerates ribulose 5-phosphate. The alternative to a transaldolase-driven rearrangement step is the use of a sedoheptulose bisphosphatase, which was not included in the initial annotation. Strøm et al. (1974) could detect no specific activity using cell-free preparations. Yet, we decided to add a corresponding reaction for two reasons. First, the FBA has been characterized to reversibly catalyze sedoheptulose bisphosphate cleavage (Rozova, Khmelenina, Mustakhimov, Reshetnikov, & Trotsenko, 2010) which is reflected by the reaction FBA3 in the model. Second, the pyrophosphate-dependent 6-phosphofructokinase (PFK_3_ppi) was reported to have higher affinity and activity to the reversible phosphorylation of seduheptulose phosphate than compared to fructose 6-phosphate (Reshetnikov et al., 2008). Thus, all of the resulting four combinations that make up the RuMP-pathway are represented in this metabolic model (Figure 4).

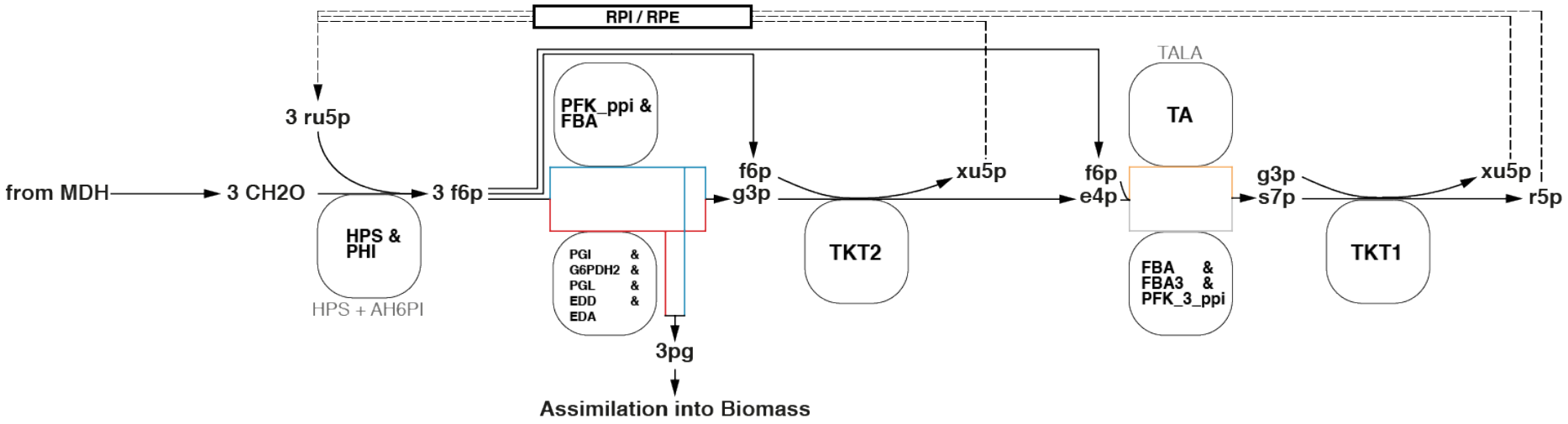


Figure 5-4. All four variants of the RuMP pathway are represented in the metabolic model. If there is a common enzyme name the black text denotes the common name, while faint grey text denotes the reaction ID in the metabolic model. Otherwise the black text is also the reaction ID in the model.

The genome of *Methylococcus capsulatus* is equipped with a complete Calvin Benson Bassham (CBB) cycle (Baxter et al., 2002; S. Taylor, 1977; Stephen Taylor, Dalton, & Dow, 1980) and a partial Serine pathway for formaldehyde assimilation (Ward et al., 2004). It was argued by Taylor et al. (1981) that *M. capsulatus can metabolize* glycolate (a product of the oxygenase activity of the ribulose bisphosphate carboxylase (RBPC)) via this pathway. Both Taylor and Ward, further suggested the presence of unique key enzymes to complete the Serine pathway, such as hydroxymethyl transferase, hydroxypyruvate reductase and malate-CoA lyase. However, since the gene annotation did not reflect this and the RuMP pathway is reportedly the main pathway for formaldehyde assimilation (C. Anthony, 1983; Kelly, Anthony, & Murrel, 2005; Strøm et al., 1974), these putative reactions were not included.

All genes of the TCA cycle were identified in the genomesequence and all associated reactions were curated accordingly (Ward et al., 2004). Because no activity of the 2-oxoglutarate dehydrogenase has so far been measured *in vivo (Kelly et al., 2005; Wood, Aurikko, & Kelly, 2004)*, the associated reactions have been constrained to be blocked (both lower and upper bounds were set to zero). This way they can easily be activated if needed. For instance, if a growth condition is discovered where activity for this enzyme can be detected.

Based on reactions already present in BMID000000141026, the information in the genome annotation, and the measured biomass composition, we curated the biosynthetic pathways of all proteinogenic amino acids, nucleotides, fatty acids, phospholipids, panthothenate, coenzyme A, NAD, FAD, FMN, riboflavin, thiamine, myo-inositol, heme, folate, cobalamine, glutathione, squalene, lanosterol, peptidoglycan. Since no corresponding genes could be identified, reactions catalyzing the biosynthesis of lipopolysaccharide (LPS) were adopted from iJO1366 (Orth et al., 2011) under the assumption that the biosynthesis steps are energetically similar to *Escherichia coli.*

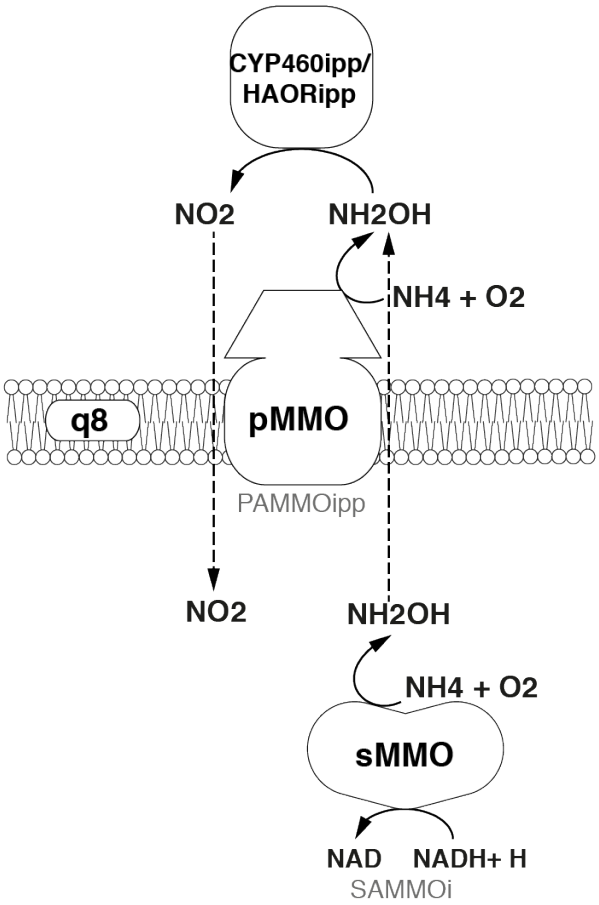


Figure 5-5. Oxidation of ammonia to nitrite.

Methylococcus capsulatus is able to metabolize the nitrogen sources ammonium (NH_4) and nitrate (NO_3) in a variety of ways. When the extracellular concentration of NH_4 is high, the alanine dehydrogenase (ADH) is the primary pathway for assimilation into biomass (Murrell & Dalton, 1983). In addition, the two monooxygenases are able to oxidize ammonium to hydroxylamine (Bédard & Knowles, 1989; Colby & Dalton, 1978), which is then further oxidized by specific enzymes first to nitrite (Bergmann, Zahn, Hooper, & DiSpirito, 1998) (Figure 5), and even to dinitrogen oxide (Campbell et al., 2011). NO_3 is reduced to NH_4 via nitrite and ultimately assimilated via the glutamine synthetase/ glutamine synthase (GS/GOGAT) pathway. Furthermore, it was shown that *M. capsulatus* is able to fix atmospheric nitrogen (N_2) (Colin Murrell & Dalton, 1983). The nitrogenase

gene cluster has been identified (Oakley & Murrell, 1991) and annotated accordingly (Ward et al., 2004), which is why the corresponding reactions have been included in the model. As the enzyme has not yet been specifically characterized, the nitrogenase reaction was adapted from iAF987 (Feist et al., 2014). A schema showing these reactions side-by-side is displayed in Figure 6.

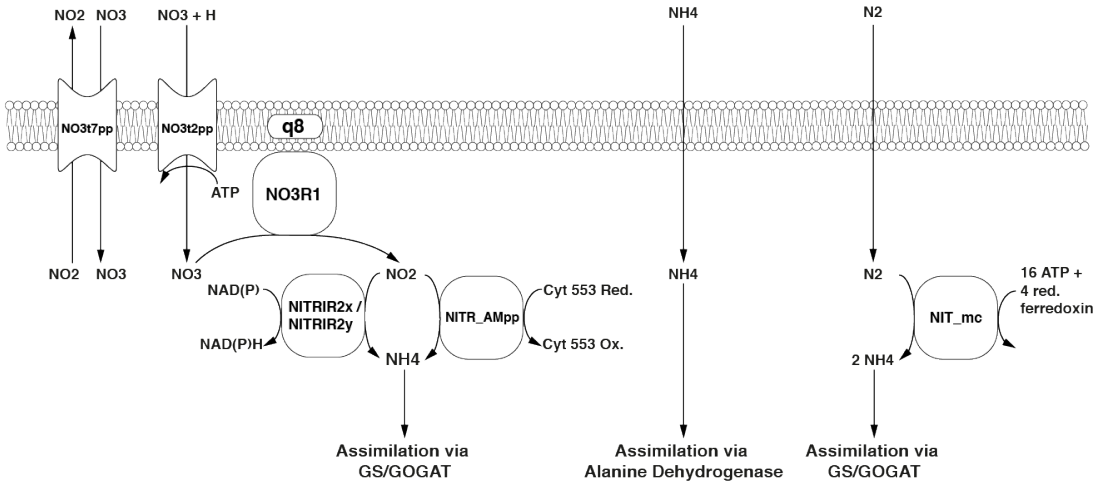


Figure 5-6. Nitrogen assimilation and fixation pathways represented in the model. Black text in the centre of the symbols denotes reaction IDs. Common names are used for metabolites.

A metabolic map of all the reactions in the model was built using Escher (King et al., 2015) and is available in the Supplement Figure 1.

5.3.4. Extension and manual curation

Starting reconstruction on the basis of an automated draft required additional effort to facilitate operability. The automated draft used two sets of ID namespaces, BiGG (King et al., 2016) and MetaNetX (MNX) (Ganter, Bernard, Moretti, Stelling, & Pagni, 2013). Hence, the first step in the curation efforts consisted of unifying the namespace by mapping all metabolite and reaction identifiers from MNX into the human-readable BiGG namespace. Individual metabolites and reactions with unmappable IDs, that could not be identified in the BiGG database and for which there was little evidence in the form of GPR rules, were removed from the model. Several metabolite formulas contained

repeating units, and many reactions lacked EC numbers. Using the MNX spreadsheet exports 'chem_prop.tsv' and 'reac_prop.tsv' from version 1.0 (Bernard et al., 2014) and 2.0 (Moretti et al., 2016) these issues were predominantly resolved. Due to said malformed and missing metabolite formulae, many reactions were not mass-charge-balanced. We used the `check_mass_balance` function in cobrapy (Ebrahim et al., 2013) to identify and balance 99% of the reactions in the model.

Another challenge with extending the automated draft was that many reactions were inverted meaning that reactants and products were switched, and constrained to be irreversible. Consulting the corresponding reactions in MetaCyc (Caspi et al., 2014), these instances were corrected manually. Following the requirements for precursors set in the biomass reaction, the corresponding biosynthetic pathways were gap-filled and manually curated top-down, using MetaCyc pathways from the *M. capsulatus* Bath specific database as a scaffold (https://biocyc.org/organism-summary?object=MCAP243233).

In order to enable other researchers to integrate iCL730 into their respective workflows and to simplify model cross-comparisons, we included MIRIAM-compliant annotations for a majority of the model's reactions (96%) and metabolites (98%).

As a last step we added transport reactions that were not in the draft reconstruction. Inferring membrane transport reactions from the genome sequence is difficult, as usually the precise 3D structure of transport proteins dictates which metabolite classes can be transported. Even if the substrates are known, the energy requirements of transport are often undefined. Working on protein sequence matches using PsortB 3.0 (Yu et al., 2010), combined with BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990) matches against TransportDB (Elbourne, Tetu, Hassan, & Paulsen, 2017) and the Transporter Classification Database (TCDB) (Saier, Tran, & Barabote, 2006), we we're able to identify 56 additional transport reactions. We have limited the amount of transporters to be added, by focussing specifically on transporters where the

mechanism was clear and which would transport metabolites that were already described in the model. A list of putative transport-associated genes that we did not consider is available at https://github.com/ChristianLieven/memote-m-capsulatus. Some of these genes are reported by Ward et al. (2004) to facilitate the uptake of sugars. Kelly et al. (2005) relate the presence of these genes to *Nitrosomas europaea,* which is able grow on fructose as a carbon-source with energy from ammonia oxidation (Hommes et al. 2003). This list may be a good starting point to study potential alternate carbon sources in *M. capsulatus*.

5.3.5. Validation of the Model

To determine which combination of the three aforementioned electron transfer modes is active in *Methylococcus capsulatus*, we constrained the model based on experiments conducted by (Leak & Dalton, 1986a). Since the three modes relate to how the pMMO receives electrons, we focused on the data generated by growing *M. capsulatus* in high-copper medium, which is the condition in which the pMMO is predominantly active. We used the average of carbon and oxygen-limited measurements as a reference. Having constrained the model, we compared the Leak and Dalton's measurements for the ratio of oxygen consumption to methane consumption (O_2/CH_4) to the predictions of the model (Figure 5-7A). We considered the O_2/CH_4 ratio to be a key metric for the respiratory chain in *M. capsulatus*, as it is a function of the mode of electron transfer to the pMMO. The central carbon metabolism was left unconstrained.

Under the assumption that the mode of electron transfer to the pMMO would be independent of the source of nitrogen, we compared the O_2/CH_4 ratios of the model constrained to employ one of the three modes of electron transfer exclusively to the corresponding reference values of *M. capsulatus* grown on either NO_3 or NH_4. However, neither of modes adequately represented the measured O_2/CH_4 ratios of about 1.43 and 1.6, respectively. Although Leak and Dalton had proposed that the *reverse* or *uphill electron transfer* is the most probable mode (Leak & Dalton, 1986b), the model predictions allowing for an unbounded *uphill transfer* did

not support this (Figure 5-7), as the efficiency was almost comparable to predictions using *direct coupling*.

To determine whether decreasing the efficiency could improve the fit, we iteratively constrained the lower bound of the reaction associated with the CYOR_q8 (CYOR_q8ipp). As Figure 5-7B shows, by constraining the reverse flux through this reaction, it is possible to modulate the ratio of O_2/CH_4 consumption. We decided to fit the lower bound of CYOR_q8 only to the reference O_2/CH_4 ratio of 1.43 with cells grown on NO_3, in order to avoid an overlap with the effects of NO_2 production. Leak and Dalton pointed out, that the unexpectedly high ratio of 1.6 proved to be due to latent NH_4 oxidation rather than assimilation (Leak & Dalton, 1986a) and thus elevated levels of NO_2. They were uncertain whether this increase could be attributed to the energetic burden of reducing NH_4 or the uncoupling effect of NO_2.

To investigate this effect, we introduced a ratio constraint (see material & methods) coupling the uptake of NH_4 to the excretion of NO_2 and explored a number of ratios (Figure 5-7C). According to the simulations, the energy spent on reducing about 50% of incoming NH_4 to NO_2 is sufficient to account for the observed, high O_2/CH_4 ratio of 1.6. Although this shows that the loss of energy could be significant enough to account for the increased ratio, this does not exclude a potential combined effect because of energy decoupling. Regardless it shows, that predictions using the metabolic model can accurately reflect the *in vivo* behavior of *M. capsulatus*.

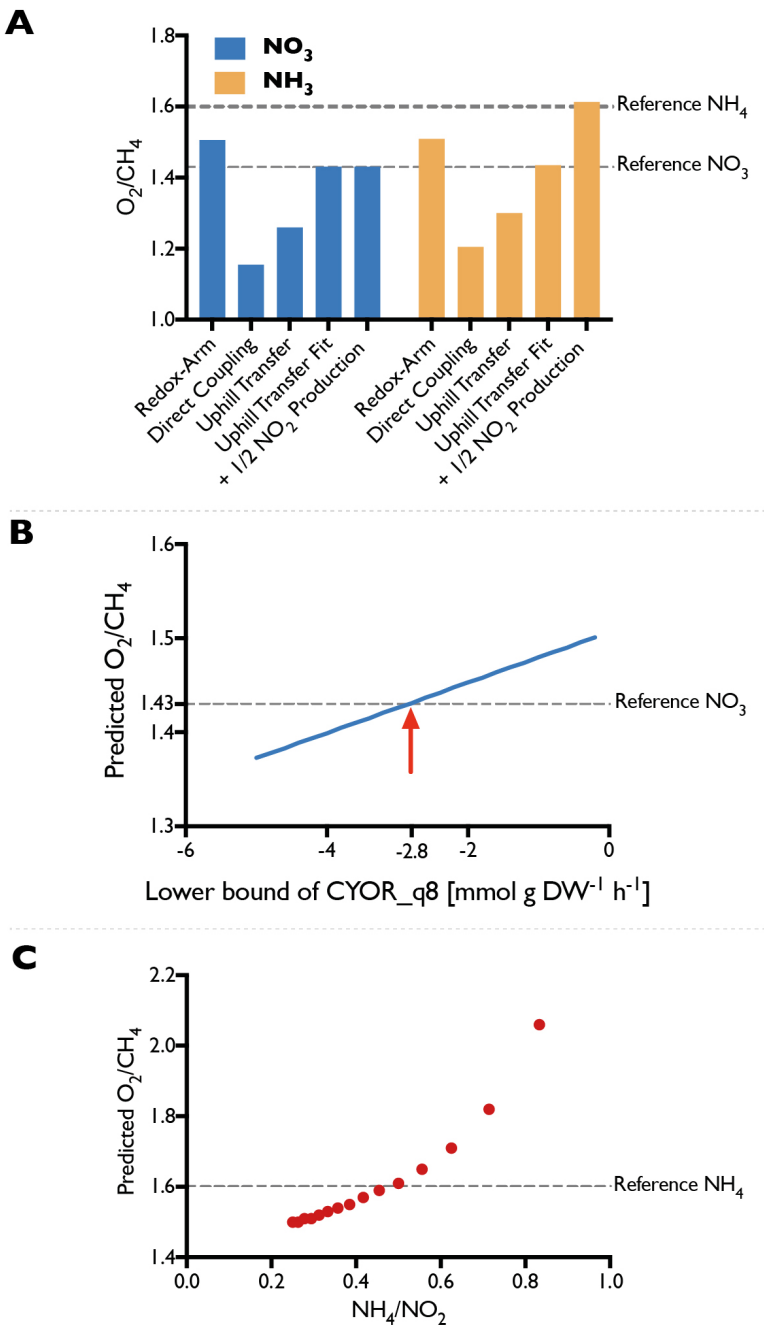


Figure 5-7 Validation of iCL730. A: Using each of the three electron transfer modes exclusively (Redox-Arm, Direct Coupling, Uphill Transfer), the ratios of oxygen to methane consumption (O_2/CH_4) predicted by iCL730 were compared to experimental values from Leak & Dalton (1986a). Since none of the three modes matched the reference, the efficiency of the Uphill Transfer was reduced by iteratively constraining the flux through the ubiquinol-cytochrome c reductase (CYOR_q8) reaction as shown in **B.** Using a lower bound of -2.8 improved the fit to the reference value for cells grown on NO_3 (**A** – Uphill Transfer Fit) **C:** To account for the energy loss through NH_3 oxidation, several ratios of NH_4 uptake to NO_2 production were considered. The closest fit was achieved with a ratio of around 0.5 (**A** – + ½ NO_2 Production)

5.3.6. Comparison with other models

A direct comparison between the here presented iCL730 with the preceding automated draft reconstruction BMID000000141026 and with a genome-scale metabolic model of *Methylomicrobium buryatense* strain 5G(B1) serves to illustrate how much the model has progressed from the draft through manual curation and how diverse Group I methanotrophs are metabolically.

Unsurprisingly, the automated draft generally performs quite poorly in comparison to the curated models. It's not possible to produce biomass from methane, even in rich media conditions, which is indicative of gaps in the pathways leading towards individual biomass components. Moreover, ATP can be produced without the consumption of substrate, which in turn means that key energy reactions are not constrained thermodynamically. In the draft, 51% of the reactions are not supported by Gene-Protein-Reaction rules (GPR), while in iMb5G(B1) it is only 32% and in iCL730 a little under 20%. GPR allow modelers to distinguish between isozymes and enzyme complexes by connecting gene IDs either through boolean 'OR' or 'AND' rules. In iCL730, 25 complexes in total were curated and formulated as GPR. Neither the draft nor iMb5G(B1) make this distinction.

In the automated draft, the oxidation of methane was only possible through a reaction representing the sMMO (MNXR6057). In iCL730, this was corrected by also implementing reactions that represent the pMMO, one with ubiquinone as the electron donor (PMMOipp), and another that receives electrons from the MDH (PMMODCipp). *Methylococcus capsulatus* is unique in that depending on the availability of copper in the medium it expresses both the soluble and the particulate monooxygenase, whereas *Methylomicrobium buryatense* only expresses the latter. In iMb5G(B1) this is represented by the reaction "pMMO". Experimental studies have thus far characterized *M. capsulatus*' ability to grow on ammonia, nitrate and nitrogen. In addition to that, however, iCL730 also predicts growth on nitrite, nitrate and urea, likely because of improperly constrained reactions. *M. buryatense* 5G(B1) is reported to grow on nitrate, ammonia and urea, yet without

adding the respective transport reactions iMb5G(B1) only grows on nitrate. While the draft contains exchange reactions for all the tested nitrogen sources except for urea, it couldn't grow at all, which again is likely because of gaps in the pathways leading to biomass precursor metabolites.

The diversity within Group 1 methanotrophs becomes even more apparent when comparing the growth energetics of iCL730 and iMb5G(B1). It is evident that *M. buryatense* is more energy efficient than *M. capsulatus*, as both on ammonia and nitrate as nitrogen sources iMb5G(B1) produces more mmol gDW^{-1} h^{-1} ATP than iCL730. Likewise, the predicted growth rate is higher in both conditions. This difference is likely a direct consequence of the mode of electron transfer to the monooxygenase and thus the efficiency of the methane oxidation reaction in total. When comparing the ratio of the uptake rate of oxygen and the uptake rate of methane for the two models, we can see that the resulting values in iMb5G(B1) are lower than in iCL730. It was recently reported, that instead of the reverse-electron transfer and redox-arm mode active in *M. capsulatus,* a mixture of *direct coupling* from pMMO to MDH and *reverse electron transfer* seems to be the most likely mode in *M. buryatense* 5G(B1) (de la Torre et al. 2015).

Table 5-1. Model Comparison. The presented reconstruction iCL730, the automated draft BMID000000141026 and iMb5G(B1), a genome-scale reconstruction of *Methylomicrobium buryatense* strain 5G(B1).

	iCL730	**BMID000000141026**	**iMb5G(B1)**
Structure			
Methane Oxidation	SMMOi, PMMOipp, PMMODCipp	MNXR6057	pMMO
Growth on N-Sources	Urea, NO_2, NO_3, NH_4, N_2	No Growth	NO_3
Performance			
ATP Production Closed Exchanges	False	True	False
ATP Production Rate - NH_4	0.775	1000	1
ATP Production Rate - NO_3	0.365	1000	0.519
Growth Rate - pMMO - NH_4	0.299	No Growth	0.34
Growth Rate - pMMO - NO_3	0.205	No Growth	0.27

O_2/CH_4 Ratio - pMMO - NH_4	1.434	No Growth	1.156
O_2/CH_4 Ratio - pMMO - NO_3	1.415	No Growth	1.116
Specifications			
Reactions without GPR	174	950	129
Enzyme Complexes	25	0	0
Total # Genes	730	589	313
Total # Metabolites	877	935	403
Total # Reactions	898	1858	402

5.4. Conclusion

ICL730 is the first, manually curated, genome-scale metabolic model for *Methylococcus capsulatus*. With iCL730, we combine biochemical knowledge of half a century of research on *Methylococcus capsulatus* into a single powerful resource, providing the basis for targeted metabolic engineering, omic-data contextualization and comparative analyses. We applied the metabolic model to study the complex electron transfer chain of *M. capsulatus*, by analyzing the three modes of electron transfer that had been proposed previously (Leak and Dalton 1986). We did so by corresponding each mode with the flux through a reaction in the model, and consequently comparing the predicted O_2/CH_4 ratios to experimentally measured values by Leak & Dalton, (1986a). Simulation and experiment agreed only when the model was constrained to employ the *uphill electron transfer* at reduced efficiency for *M. capsulatus* grown on NO_3 as the source of nitrogen. The experimentally observed effect of NH_4 oxidation to NO_2 could be replicated by considering the energy burden alone.

Future applications of the metabolic model could include hypothesis testing of the regulation of the MMO in other growth conditions (Kelly et al. 2005), studying the effects of the predicted hydrogenases on the energy metabolism of *M. capsulatus* (Ward et al. 2004), and exploring venues of metabolic engineering for an improved production of metabolites (Kalyuzhnaya et al. 2015).

5.5. Material and Methods

5.5.1. Model Curation

After mapping the reaction and metabolite identifiers from the MetaNetX namespace to the BiGG namespace, we proceeded with the curation efforts as follows: First, we chose a subsystem of interest, then we picked a pathway and using information from either the genome sequence, published articles, the metacyc or uniprot databases, and lastly, we enriched each enzymatic step in said pathway with as much information as possible. Information that was added included for instance: GPR, reaction localization, EC numbers, a confidence score, possible cofactors and inhibitors and cross references to other databases such as KEGG, BIGG and MetaNetX. For each metabolite involved in these reactions, we defined the stoichiometry, charge and elemental formula, either based on the corresponding entries in the BiGG database or on clues from literature.

If reactions from a pathway were present in the draft, we checked their constraints and directionality. This was necessary as many of the irreversible reactions in the draft reconstruction seemed to have been 'inverted' when compared to the corresponding reactions in the reference databases, which made flux through them impossible in normal growth conditions.

The energy metabolism and methane oxidation were curated first. Except for the reaction representing the sMMO, all reactions were newly constructed, as they were absent in the draft. Then, in order to achieve sensible FBA solutions for growth on methane, the central carbon metabolism, amino acid and nucleotide biosynthesis pathways were manually curated. Simultaneous to the manual curation a metabolic pathway map was created in Escher, which helped us to maintain a visual checklist of curated pathways.

The automated draft contained a rudimentary, generic biomass reaction, which only accounted for the production of proteins, DNA and RNA, but not for the

biosynthesis of a cell wall and cell membrane, the maintenance of a cofactor pool, the turnover of trace elements or the energetic costs associated with growth. After calculating a more specific biomass composition for *M. capsulatus*, further pathway curation was necessary to achieve growth *in silico*. This included the biosynthesis pathways of fatty acids (even, odd and cyclopropane varieties), phospholipids, coenzyme A, Ubiquinol, Lanosterol, FAD, FMN, Riboflavin, NAD and NADP, Heme, Glutathione, Cobalamin, Thiamine, Myo-Inositol, and Lipopolysaccharides.

To account for the reported differences in ammonia assimilation of *M. capsulatus* when grown in the presence of excess ammonia versus the growth on either atmospheric nitrogen or nitrate, we curated the nitrogen metabolism including the oxidation of ammonia to nitrite via hydroxylamine, the reduction of nitrate and nitrite, the glutamine synthetase/ glutamate synthase reactions and the alanine dehydrogenase. Reversible degradation reactions producing ammonia that would potentially bypass the characterized ammonia assimilation pathways were constrained to be irreversible accordingly.

After we had enriched the annotations already in the draft with annotations from the metabolic models iJO1366 (Orth et al., 2011), iRC1080 (Chang et al., 2011), iJN678 (Nogales, Gudmundsson, Knight, Palsson, & Thiele, 2012) and iHN637 (Nagarajan et al., 2013), they were converted into a MIRIAM-compliant format. As a final step, we manually added transport reactions to reflect the uptake energetics of cofactors.

Throughout the reconstruction process, we iteratively tested and validated the reconstruction. For instance, we checked the mass and charge balance of each reaction, attempting to manually balance those that weren't balanced. In the majority of cases metabolites were missing formula or charge definitions. In order to remove energy generating cycles, problematic reactions were manually constrained to be irreversible. Validation was carried out against growth data from (Leak & Dalton, 1986b), which was also the point of reference for the parameter fitting.

5.5.2. Biomass Composition

For the principal components of biomass, measurements were made available by our collaborators UniBio A/S (http://www.unibio.dk/end-product/chemical-composition-1). This included specifically the content of RNA (6.7%), DNA (2.3%), crude fat (9.1%), and glucose (4.5%) as a percentage of the cell dry weight. We did not use the percentage values for crude protein (72.9%) and N-free extracts (7.6%) as these measurements are fairly inaccurate relying on very generalized factors. The percentage value of Ash 550 (8.6%) measurements was inconsistent with the sum of its individual components (4.6%) and was hence excluded.

On the homepage of UniBio A/S, we were also able to find g/kg measurements of all amino acids except for glutamine and asparagine, trace elements and vitamins, which could directly be converted into mmol/g DW. However, we omitted some of data: The stoichiometries for Selenium, Cadmium, Arsenic, Lead and Mercury were not included in the biomass reaction as their values were negligibly small. Beta-Carotene (Vitamin A) and Gama-Tocopherol (Vitamin E) were omitted because no genes were found supporting their biosynthesis, in addition to both being reportedly below the detection limit (M. Øverland, Schøyen, & Skrede, 2010).

For the lack of better measurements, and assuming that *M. buryatense* and *M. capsulatus* are more similar than *M. capsulatus* and *E. coli,* the stoichiometries of glutamine and asparagine, intracellular metabolites such as ribulose-5-phosphate, organic cofactors such as coenzyme A, and cell wall components such as LPS were introduced from (de la Torre et al., 2015).

Using the GC content calculated from the genome sequence (Ward et al., 2004) and the percentage measurements from UniBio for RNA and DNA, we were able to calculate the stoichiometries of all individual nucleobases.

UniBio's measurements that 94% of the crude fat portion were fatty acids conflicted with previously published results (Makula, 1978; Müller, Hellgren, Olsen, & Skrede, 2004), which indicated that in fact phospholipids are likely to be the main lipid component in *M. capsulatus*. Thus, we assumed 94% of the crude fat to be phospholipids. This meant that 6% of the crude fat was composed of fatty acids, the distributions of which were again provided by UniBio. However, in order to calculate the stoichiometry of each fatty acid we recalculated the distribution to exclude the unidentified part. (Makula, 1978) had also measured the composition of the phospholipid pool itself, from which we calculated the corresponding stoichiometries for phosphatidylethanolamine, phosphatidylcholine, phosphatidylglycerol and cardiolipin.

(Bird et al., 1971) had reported the percentage of squalene and sterols of dry weight, which we converted into mmol/g DW without further corrections. Since the genes for hopanoid synthesis were identified (Nakano, Motegi, Sato, Onodera, & Hoshino, 2007; Ward et al., 2004) we included diplopterol in the biomass reaction. For a lack of more detailed measurements we estimated a similar contribution to the overall composition as squalene. We specifically used lanosterol to represent the general class of sterols in the biomass reaction, since we had only been able to reconstruct the synthesis pathway of lanosterol and since lanosterol is a key precursor metabolite in the biosynthesis of all other sterols.

The growth associated maintenance energy requirements were calculated in accordance with the procedures outlined by (Thiele & Palsson, 2010).

5.5.3. Transport Reactions

The identification of transport reactions involved the two databases for transport proteins, the Transporter Classification Database (TCDB) (Saier et al., 2006) and the Transport Database (TransportDB) (Elbourne et al., 2017), and two computational tools, PSORTb v3.0 (Yu et al., 2010) and BLASTp (Altschul et al., 1990). We employed the following semi-automatic way of inferring the putative function of transport proteins in *M. capsulatus*.

Using the protein sequences in AE017282.2_protein.faa (Ward et al., 2004), we determined the subcellular location of each protein using PSORTb v3.0. We filtered the results and focused only on proteins with a final score larger than 7.5, which the authors of PSORTb consider to be meaningful. We combined this list with the *M. capsulatus* specific entries from the TransportDB, which allowed us to use the PSORT-scores as an additional measure of confidence. At this point, 242 putative transport proteins were identified. From this list we then selected all proteins which were predicted to transport metabolites and were already included in the model, which reduced the number to 133. Since for many of the entries, the exact mechanism of transport is unknown, we combined the previously selected transporters with the results from running BLAST against the TCDB. The e-value and bitscore from BLAST provided yet another measure to confidently assess the correctness of the automatic TransportDB predictions, and the Transporter Classification-Numbers (TC-numbers) allowed us to gather information on the mechanism of transport. This led to a list of 97 transport proteins with high confidence, which was filtered once more as follows.

We checked the general description for a given specific TC-number, and then considered the BLAST result to read about a given transporter in more detail, especially with regards to finding the specific substrates. When we were able to identify the corresponding transport reaction in the BiGG database, we incorporated only the simplest, smallest set of reactions. In cases of conflicts between our own BLAST results and the automatic TransportDB transporter annotation, we preferentially trusted the BLAST results. Thus we ultimately added 75 transporter-encoding genes connected via GPR to 56 distinct transport reactions.

5.5.4. Ratio Constraint

To identify which mode of electron transfer is active in *M. capsulatus*, we fit the solutions of FBA using iCL730 to measured values from (Leak & Dalton, 1986a). The authors had experimentally determined the O_2/CH_4 ratio and the growth yield of *M. capsulatus* in several conditions. They varied the nitrogen source using

KNO_3, NH_4CL, and both simultaneously; the concentration of copper in the medium, which directly affects the activity of either sMMO or pMMO; and whether the culture was oxygen or carbon limited.

Since each electron transfer mode respectively is represented by the flux through one specific reaction in the model (PMMOipp, PMMODCipp, CYOR_q8), we were able to investigate each simply by constraining the corresponding reaction.

Secondly, we accounted for the differential expression of ADH in the presence of excess NH_4 in the medium versus the GS/GOGAT in the presence of NO_4 or N_2 (Murrell & Dalton, 1983) by blocking the corresponding reactions accordingly.

Several studies have shown that NH_4 is a co-metabolic substrate to the methane monooxygenases in *M. capsulatus* leading to the production of hydroxylamine first and nitrite later (Bédard & Knowles, 1989; Hutton & Zobell, 1953; Nyerges & Stein, 2009). Hence, when simulating the growth on NH_4 we assumed that varying ratios $\boldsymbol{r}$ of the nitrogen that was taken up would eventually be converted into nitrite (Figure 5-7).

(7) $$\boldsymbol{v_{EX\_nh4\_e} + r v_{EX\_no2\_e} = 0}$$

5.5.5. Stoichiometric Modeling

The reactome of an organism can be represented mathematically as a stoichiometric matrix $\boldsymbol{S}$. Each row of $\boldsymbol{S}$ corresponds to a unique compound, while each column corresponds to a metabolic reaction. Hence the structure of the matrix for an organism with $\boldsymbol{m}$ compounds and $\boldsymbol{n}$ reactions equals $\boldsymbol{m \times n}$. The values in each row denote the stoichiometric coefficients of that metabolite in all reactions. The coefficients are either negative for compounds that are educts, positive for those that are products, or zero for those that are not involved in a given metabolic reaction.

The vector $\boldsymbol{v}$ of length $\boldsymbol{n}$ contains as values the turnover rates of each reaction. These rates are commonly referred to as fluxes and are given the unit mmol gDW^{-1} h^{-1}. Vector $\boldsymbol{v}$ is also referred to as the flux vector,

5.6. Acknowledgements

The authors would like to acknowledge Leander Petersen, Sten Bay Jorgensen, Budi Juliman Hidayat, Subir Kumar Nandi, John Villadsen for the countless hours of discussions about their experiences with M. capsulatus; Mitch Pesesky, David Collins, Marina Kalyuzhnaya, Ilya Akberdin, and Sergey Stolyar for insightful discussions at the GRC C1 Conference 2016; and Kristian Jensen, Joao Cardoso, and Marta Matos for their help with software problems and feedback on the implementation. This work has been funded by the Novo Nordisk Foundation and the Innovation Fund Denmark (project "Environmentally Friendly Protein Production (EFPro2)")

5.7. Supplement

The metabolic model iCL730 and scripts that were used to construct it, in addition to the scripts that were used for the analysis presented in this work, are provided at https://github.com/ChristianLieven/memote-m-capsulatus.

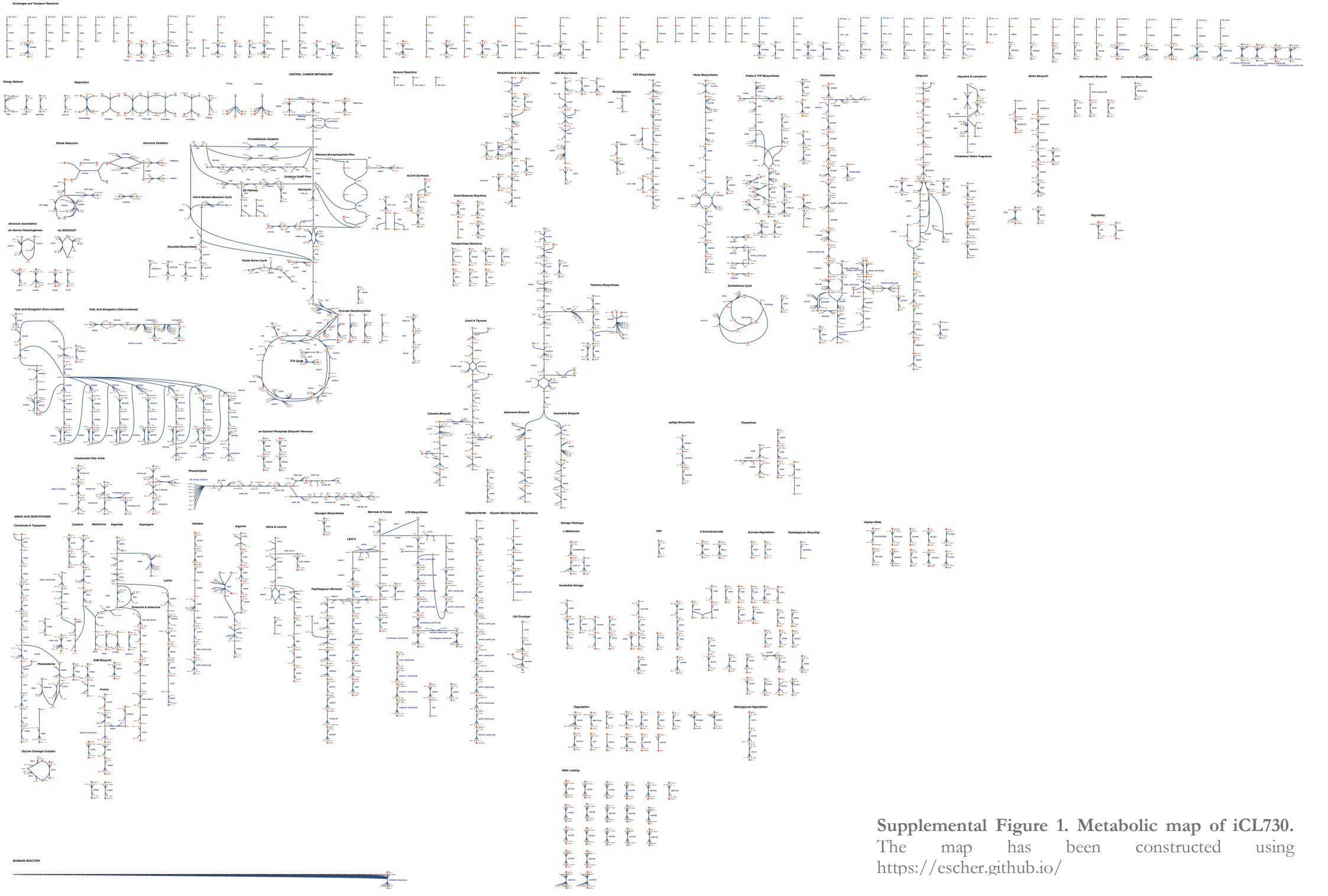
Supplemental Table 1: Stoichiometry of the biomass reaction in iCL730.

Compound	mmol $gDCW^{-1}$	References
Proteins		
Ala	0.576	unibio.dk[9], Øverland et al, 2010[10]
Arg	0.254	"
Asp	0.468	"
Cys	0.038	"
Glu	0.519	"
Gly	0.484	"

	Compound	mmol gDCW^{-1}	References
	His	0.104	‘’
	Ile	0.242	‘’
	Leu	0.415	‘’
	Lys	0.277	‘’
	Met	0.130	‘’
	Phe	0.185	‘’
	Pro	0.248	‘’
	Ser	0.246	‘’
	Thr	0.271	‘’
	Trp	0.093	‘’
	Tyr	0.132	‘’
	Val	0.360	‘’
	Gln	0.150	de la Torre et al, 2015[4]
	Asn	0.119	‘’
Nucleic Acids	*Ribonucleic acid (RNA)*		
	A	0.024	unibio.dk; Ward et al,2004[11]
	U	0.026	‘’
	G	0.041	‘’
	C	0.044	‘’
	Deoxyribonucleic acid (DNA)		
	A	0.009	‘’
	T	0.009	‘’
	G	0.014	‘’
	C	0.016	‘’
Lipids	*Phospholipids*		
	PE	0.081	Average of 4 experiments by Makula 1978, converted from measurements in μmol/g DW[12]
	PG	0.014	‘’
	PC	0.008	‘’
	CL	0.003	‘’
	Sterols		
	Squalene	0.013	Bird et al. ,1971[13]
	Diplopterol	0.013	Estimated
	Lanosterol	0.005	Bird et al. ,1971; Zymosterol was detected
	Fatty acids		
	C14:0 Myristic acid	0.001	unibio.dk, Müller et al, 2004[14]
	C14:1	0.000	‘’
	C15:0	0.000	‘’

	Compound	mmol gDCW^{-1}	References
	C16:0 Palmitic acid	0.016	‘’
	C17:00	0.002	‘’
	Cyc17:0	0.002	‘’
	C18:0 Stearic acid	0.000	‘’
	C18:1 Oleic acid	0.000	‘’
Salts			
	Phosphorous	0.200	unibio.dk
	Sulphur	0.056	‘’
	Chloride	0.214	‘’
	Calcium	0.070	‘’
	Potassium	0.176	‘’
	Magnesium	0.123	‘’
	Sodium	0.039	‘’
	Iron	0.006	‘’
	Copper	0.002	‘’
	Zinc	0.001	‘’
	Cobalt	0.000	‘’
	Nickel	0.000	‘’
	Manganese	0.000	‘’
Cell Wall			
	Peptidoglycan	0.053	‘’
	LPS	0.002	‘’
	Glucose	0.250	‘’
Vitamins			
	Thiamin B1	0.000	‘’
	Riboflavin B2	0.000	‘’
	Nicotinic acid	0.001	‘’
	Inositol 30	0.000	‘’
Intracellular Metabolites			
	Ribulose 5-phospate	0.001	de la Torre et al, 2015
	Fructose-1,6-bisphosphate	0.001	‘’
	Fructose-6-phospate	0.003	‘’
	Glucose-6-phosphate	0.002	‘’
	Glyceraldehyde-3-phosphate/ dhap	0.003	‘’
	6-phospogluconic acid	0.000	‘’
	2-dehydro-3-deoxy-phosphogluconate	0.000	‘’
	Phosphoglycerate	0.006	‘’
	Phosphoenolpyruvate	0.005	‘’
	Pyruvate	0.015	‘’
	Acetyl-CoA	0.000	‘’
	Succinate	0.002	‘’
	Malate	0.004	‘’
	Fumarate	0.001	‘’

Compound	mmol gDCW^{-1}	References
Citrate	0.001	‘’
Glycerate	0.001	‘’
Nitrate	0.002	unibio.dk
Nitrite	0.002	‘’
Energy Requirement		
ADP	-23.087	Calculated from Thiele and Palsson, 2010
Pi	-23.087	‘’
Ppi	-0.183	‘’
H	-23.087	‘’
H2O	17.776	‘’
ATP	23.087	‘’



Supplemental Figure 1. Metabolic map of iCL730. The map has been constructed using https://escher.github.io/

5.8. References

Adeosun, E. K., Smith, T. J., Hoberg, A. M., Velarde, G., Ford, R., & Dalton, H. (2004). Formaldehyde dehydrogenase preparations from Methylococcus capsulatus (Bath) comprise methanol dehydrogenase and methylene tetrahydromethanopterin dehydrogenase. *Microbiology*, *150*(3), 707–713.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Anthony, C. (1983). *The biochemistry of methylotrophs* (Vol. 75, p. 497).

Anthony, C. (1992). The c-type cytochromes of methylotrophic bacteria. *Biochimica et Biophysica Acta*, *1099*, 1–15.

Baxter, N. J., Hirt, R. P., Bodrossy, L., Kovacs, K. L., Embley, T. M., Prosser, J. I., & Murrell, J. C. (2002). The ribulose-1,5-bisphosphate carboxylase/oxygenase gene cluster of Methylococcus capsulatus (Bath). *Archives of Microbiology*, *177*(4), 279–289.

Bédard, C., & Knowles, R. (1989). Physiology, biochemistry, and specific inhibitors of CH4, NH4+, and CO oxidation by methanotrophs and nitrifiers. *Microbiological Reviews*, *53*(1), 68–84.

Bergmann, D. J., Zahn, J. a., Hooper, A. B., & DiSpirito, A. a. (1998). Cytochrome P460 genes from the methanotroph Methylococcus capsulatus bath. *Journal of Bacteriology*, *180*(24), 6440–6445.

Bernard, T., Bridge, A., Morgat, A., Moretti, S., Xenarios, I., & Pagni, M. (2014). Reconciliation of metabolites and biochemical reactions for metabolic networks. *Briefings in Bioinformatics*, *15*(1), 123–135.

Bird, C. W., Lynch, J. M., Pirt, F. J., Reid, W. W., Brooks, C. J. W., & Middleditch, B. S. (1971). Steroids and Squalene in Methylococcus capsulatus grown on Methane. *Nature*, *230*(5294), 473–474.

Blazyk, J. L., & Lippard, S. J. (2002). Expression and characterization of ferredoxin and flavin adenine dinucleotide binding domains of the reductase component of soluble methane monooxygenase from Methylococcus capsulatus (Bath). *Biochemistry*, *41*(52), 15780–15794.

Büchel, F., Rodriguez, N., Swainston, N., Wrzodek, C., Czauderna, T., Keller, R., … Le Novère, N. (2013). Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Systems Biology*, *7*, 116.

Campbell, M. A., Nyerges, G., Kozlowski, J. A., Poret-Peterson, A. T., Stein, L. Y., & Klotz, M. G. (2011). Model of the molecular basis for hydroxylamine oxidation and nitrous oxide production in methanotrophic bacteria. *FEMS Microbiology Letters*, *322*(1), 82–89.

Cardoso, J., Jensen, K., Lieven, C., & Hansen, A. (2017). Cameo: A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories. *bioRxiv*. Retrieved from http://biorxiv.org/content/early/2017/06/09/147199.abstract

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., … Karp, P. D. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, *42*. https://doi.org/10.1093/nar/gkt1103

Chang, R. L., Ghamsari, L., Manichaikul, A., Hom, E. F. Y., Balaji, S., Fu, W., … Papin, J. A. (2011). Metabolic network reconstruction of Chlamydomonas offers insight into light- driven algal metabolism. *Molecular Systems Biology*, *7*(1), 518.

Choi, D.-W., Kunz, R. C., Boyd, E. S., Jeremy, D., Antholine, W. E., Han, J., … Semrau, J. D. (2003). The Membrane-Associated Methane Monooxygenase (pMMO) and pMMO-NADH : Quinone Oxidoreductase Complex from Methylococcus capsulatus Bath The Membrane-Associated Methane Monooxygenase (pMMO) and pMMO-NADH : Quinone Oxidoreductase Complex from Methylococc. *Journal of Bacteriology*, *185*(19), 5755–5764.

Colby, J., & Dalton, H. (1978). Resolution of the methane mono-oxygenase of Methylococcus capsulatus (Bath) into three components. Purification and properties of component C, a flavoprotein. *Biochemical Journal*, *171*(2), 461–468.

Colby, J., & Dalton, H. (1979). Characterization of the second prosthetic group of the flavoenzyme NADH–acceptor reductase (component C) of the methane mono-oxygenase from Methylococcus capsulatus (Bath). *Biochemical Journal*, *177*(3), 903–908.

Colin Murrell, J., & Dalton, H. (1983). Nitrogen Fixation in Obligate Methanotrophs. *Microbiology*, *129*(11), 3481–3486.

Culpepper, M. A., & Rosenzweig, A. C. (2014). Structure and Protein–Protein Interactions of Methanol Dehydrogenase from Methylococcus capsulatus (Bath). *Biochemistry*, *53*(39), 6211–6219.

de la Torre, A., Metivier, A., Chu, F., Laurens, L. M. L., Beck, D. A. C., Pienkos, P. T., … Kalyuzhnaya, M. G. (2015). Genome-scale metabolic reconstructions and theoretical investigation of methane conversion in Methylomicrobium buryatense strain 5G(B1). *Microbial Cell Factories*, *14*(1), 188.

DiSpirito, A. A., Kunz, R. C., Choi, D.-W., & Zahn, J. A. (2004). Respiration in methanotrophs. In *Respiration in archaea and bacteria* (pp. 149–168). Springer.

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, *7*, 74.

Elbourne, L. D. H., Tetu, S. G., Hassan, K. A., & Paulsen, I. T. (2017). TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Research*, *45*(D1), D320–D324.

Feist, A. M., Nagarajan, H., Rotaru, A.-E., Tremblay, P.-L., Zhang, T., Nevin, K. P., … Zengler, K. (2014). Constraint-based modeling of carbon fixation and the energetics of electron transfer in Geobacter metallireducens. *PLoS Computational Biology*, *10*(4), e1003575.

Feist, A. M., Zielinski, D. C., Orth, J. D., Schellenberger, J., Herrgard, M. J., & Palsson, B. O. (2010). Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. *Metabolic Engineering*, *12*(3), 173–186.

Foster, J. W., & Davis, R. H. (1966). A methane-dependent coccus, with notes on classification and nomenclature of obligate, methane-utilizing bacteria. *Journal of Bacteriology*, *91*(5), 1924–1931.

Ganter, M., Bernard, T., Moretti, S., Stelling, J., & Pagni, M. (2013). MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics* , *29*(6), 815–816.

Hanson, R. S., & Hanson, T. E. (1996). Methanotrophic bacteria. *Microbiological Reviews*, *60*(2), 439–471.

Hutton, W. E., & Zobell, C. E. (1953). Production of nitrite from ammonia by methane oxidizing bacteria. *Journal of Bacteriology*, *65*(2), 216–219.

Indrelid, S., Kleiveland, C., Holst, R., Jacobsen, M., & Lea, T. (2017). The Soil Bacterium Methylococcus capsulatus Bath Interacts with Human Dendritic Cells to Modulate Immune Function. *Frontiers in Microbiology*, *8*(February), 1–13.

Jiang, H., Chen, Y., Jiang, P., Zhang, C., Smith, T. J., Murrell, J. C., & Xing, X.-H. (2010). Methanotrophs: Multifunctional bacteria with promising applications in environmental bioengineering. *Biochemical Engineering Journal*, *49*(3), 277–288.

Kelly, D. P., Anthony, C., & Murrel, J. C. (2005). Insights into the obligate methanotroph Methylococcus capsulatus. *Trends in Microbiology*, *13*(5), 195–198.

King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., & Palsson, B. O. (2015). Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Computational Biology*, *11*(8). https://doi.org/10.1371/journal.pcbi.1004321

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., … Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, *44*(D1), D515–22.

Kleiveland, C. R., Hult, L. T. O., Spetalen, S., Kaldhusdal, M., Christofferesen, T. E., Bengtsson, O., … Leaa, T. (2013). The noncommensal bacterium Methylococcus capsulatus (Bath) ameliorates dextran sulfate (sodium salt)-induced ulcerative colitis by influencing mechanisms essential for maintenance of the colonic barrier function. *Applied and Environmental Microbiology*, *79*(1), 48–57.

Larsen, Ø., & Karlsen, O. A. (2016). Transcriptomic profiling of Methylococcus capsulatus (Bath) during growth with two different methane monooxygenases. *MicrobiologyOpen*, *5*(2), 254–267.

Leak, D. J., & Dalton, H. (1986a). Growth yields of methanotrophs - 1. Effect of copper on the energetics of methane oxidation. *Applied Microbiology and Biotechnology*, *23*(6), 470–476.

Leak, D. J., & Dalton, H. (1986b). Growth yields of methanotrophs 2. A theoretical analysis. *Applied Microbiology and Biotechnology*, *23*(6), 477–481.

Lund, J., Woodland, M. P., & Dalton, H. (1985). Electron transfer reactions in the soluble methane monooxygenase of Methylococcus capsulatus (Bath). *European Journal of Biochemistry / FEBS*, *147*(2), 297–305.

Makula, R. A. (1978). Phospholipid composition of methane-utilizing bacteria. *Journal of Bacteriology*, *134*(3), 771–777.

Moretti, S., Martin, O., Van Du Tran, T., Bridge, A., Morgat, A., & Pagni, M. (2016). MetaNetX/MNXref--reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Research*, *44*(D1), D523–D526.

Müller, H., Hellgren, L. I., Olsen, E., & Skrede, A. (2004). Lipids rich in phosphatidylethanolamine from natural gas-utilizing bacteria reduce plasma cholesterol and classes of phospholipids: A comparison with soybean oil. *Lipids*, *39*(9), 833–841.

Murrell, J. C., & Dalton, H. (1983). Ammonia Assimilation in Methylococcus-Capsulatus (Bath) and Other Obligate Methanotrophs. *Journal of General Microbiology*. https://doi.org/10.1099/00221287-129-4-1197

Nagarajan, H., Sahin, M., Nogales, J., Latif, H., Lovley, D. R., Ebrahim, A., & Zengler, K. (2013). Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of Clostridium ljungdahlii. *Microbial Cell Factories*, *12*, 118.

Nakano, C., Motegi, A., Sato, T., Onodera, M., & Hoshino, T. (2007). Sterol biosynthesis by a prokaryote: first in vitro identification of the genes encoding squalene epoxidase and lanosterol synthase from Methylococcus capsulatus. *Bioscience, Biotechnology, and Biochemistry*, *71*(10), 2543–2550.

Neidhardt, F. C., Ingraham, J. L., & Schaechter, M. (1990). *Physiology of the bacterial cell: a molecular approach* (Vol. 20). Sinauer Sunderland.

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O., & Thiele, I. (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(7), 2678–2683.

Nunes et al., J. J. (2016). Enhanced production of single cell protein from M. capsulatus (Bath) growing in mixed culture. *Journal of Microbiology, Biotechnology and Food Sciences*, *6*(3), 894–899.

Nyerges, G., & Stein, L. Y. (2009). Ammonia cometabolism and product inhibition vary considerably among species of methanotrophic bacteria. *FEMS Microbiology Letters*, *297*(1), 131–136.

Oakley, C. J., & Murrell, J. C. (1991). Cloning of nitrogenase structural genes from the obligate methanotroph Methylococcus capsulatus (Bath). *FEMS Microbiology Letters*, *62*(2-3), 121–125.

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., & Palsson, B. Ø. (2011). A comprehensive genome- scale reconstruction of Escherichia coli metabolism—2011. *Molecular Systems Biology*, *7*(1), 535.

Øverland, M., Schøyen, H. F., & Skrede, A. (2010). Growth performance and carcase quality in broiler chickens fed on bacterial protein grown on natural gas. *British Poultry Science*, *51*(5), 686–695.

Øverland, M., Tauson, A.-H., Shearer, K., & Skrede, A. (2010). Evaluation of methane-utilising bacteria products as feed ingredients for monogastric animals. *Archives of Animal Nutrition*, *64*(3), 171–189.

Petersen, L., Villadsen - Biotechnology and …, J., & 2017. (2017). Mixing and mass transfer in a pilot scale U- loop bioreactor. *Wiley Online Library*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/bit.26084/full

Reshetnikov, A. S., Rozova, O. N., Khmelenina, V. N., Mustakhimov, I. I., Beschastny, A. P., Murrell, J. C., & Trotsenko, Y. A. (2008). Characterization of the pyrophosphate-dependent 6-phosphofructokinase from Methylococcus capsulatus Bath, *288*, 202–210.

Ritala, A., Häkkinen, S. T., Toivari, M., & Wiebe, M. G. (2017). Single Cell Protein—State-of-the-Art, Industrial Landscape and Patents 2001–2016. *Frontiers in Microbiology*, *8*, 2009.

Ross, M. O., & Rosenzweig, A. C. (2017). A tale of two methane monooxygenases. *Journal of Biological Inorganic Chemistry: JBIC: A Publication of the Society of Biological Inorganic Chemistry*, *22*(2-3), 307–319.

Rozova, O. N., Khmelenina, V. N., Mustakhimov, I. I., Reshetnikov, A. S., & Trotsenko, Y. A. (2010). Characterization of Recombinant Fructose 1 , 6 Bisphosphate Aldolase from Methylococcus capsulatus Bath, *75*(7), 892–898.

Saier, M. H., Jr, Tran, C. V., & Barabote, R. D. (2006). TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Research*, *34*(Database issue), D181–6.

Shiemke, A. K., Arp, D. J., & Sayavedra-Soto, L. A. (2004). Inhibition of membrane-bound methane monooxygenase and ammonia monooxygenase by diphenyliodonium: implications for electron transfer. *Journal of Bacteriology*, *186*(4), 928–937.

Strøm, T., Ferenci, T., & Quayle, J. R. (1974). The carbon assimilation pathways of Methylococcus capsulatus, Pseudomonas methanica and Methylosinus trichosporium (OB3B) during growth on methane. *Biochemical Journal*, *144*(3), 465–476.

Tate, S., & Dalton, H. (1999). A low-molecular-mass protein from Methylococcus capsulatus (Bath) is responsible for the regulation of formaldehyde dehydrogenase activity in vitro. *Microbiology*, *145 (Pt 1)*, 159–167.

Taylor, S. (1977). Evidence for the presence of ribulose 1,5-bisphosphate carboxylase and phosphoribonuclease in Methylococcus capsulatus (bath). *FEMS Microbiology Letters*, *2*(6), 305–307.

Taylor, S. C., Dalton, H., & Dow, C. S. (1981). Ribulose-1,5-bisphosphate Carboxylase/Oxygenase and Carbon Assimilation in Methylococcus capsulatus (Bath). *Microbiology*, *122*(1), 89–94.

Taylor, S., Dalton, H., & Dow, C. (1980). Purification and initial characterisation of ribulose 1,5-bisphosphate carboxylase from Methylococcus capsulatus (Bath). *FEMS Microbiology Letters*, *8*(3), 157–160.

Thiele, I., & Palsson, B. Ø. (2010, January). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*. Nature Publishing Group. https://doi.org/10.1038/nprot.2009.203

Trotsenko, Y. A., & Murrell, J. C. (2008). Metabolic Aspects of Aerobic Obligate Methanotrophy, *63*(07), 5–6.

Vorholt, J. A. (2002). Cofactor-dependent pathways of formaldehyde oxidation in methylotrophic bacteria. *Archives of Microbiology*, *178*(4), 239–249.

Ward, N., Larsen, Ø., Sakwa, J., Bruseth, L., Khouri, H., Durkin, a. S., … Eisen, J. a. (2004). Genomic insights into methanotrophy: The complete genome sequence of Methylococcus capsulatus (Bath). *PLoS Biology*, *2*(10). https://doi.org/10.1371/journal.pbio.0020303

Wood, A. P., Aurikko, J. P., & Kelly, D. P. (2004). A challenge for 21st century molecular biology and biochemistry: What are the causes of obligate autotrophy and methanotrophy? *FEMS Microbiology Reviews*, *28*(3), 335–352.

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., … Brinkman, F. S. L. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* , *26*(13), 1608–1615.

Zahn, J. a., Bergmann, D. J., Boyd, J. M., Kunz, R. C., & DiSpirito, A. a. (2001). Membrane-associated quinoprotein formaldehyde dehydrogenase from Methylococcus capsulatus Bath. *Journal of Bacteriology*, *183*(23), 6832–6840.

Conclusion, Future Plans and Prospects

Computer-aided design can be immensely powerful, especially to investigate phenomena that humans alone cannot understand intuitively. Metabolism and the behavior that arises from its network of interconnections is a phenomenon that becomes much more navigable with the plethora of available computational methods. Being able to generate genome-scale metabolic models from the genome sequence of any organism or even entire meta-genomes allows researchers to interrogate the mechanistic behavior of organisms that are difficult to culture, let alone isolate from their environment.

This could prove particularly useful when studying the role of yet unchartered sinks (and sources) in the global methane cycle. As outlined in Chapter 3, genome-scale metabolic models are already on the rise, as a means to not only improve our understanding of methanotrophy in isolated methanotrophs, but also to further explore their potential as industrial biocatalysts. The benefits are clear: methane is cheap and abundant; and removing it from the environment eliminates its threat to the climate as the second most prominent greenhouse gas after carbon dioxide.

Unfortunately, and despite of the recent hype, many challenges remain. Atmospheric methane itself, while being an excellent feedstock in terms of yield and biosustainability considerations, ultimately dissolves poorly in water, which requires innovative reactor technologies. In comparison to the contemporary microbial workhorses *E. coli* and *S. cerevisiae*, Methanotrophs are underexplored and robust molecular methods are scarce. This in turn means that the physiological data, required to curate and refine metabolic models, is equally limited.

While even comparably bare models can function as predictive tools merely on account of connecting genome-scale stoichiometric information, more powerful predictions are feasible only after integrating large-scale phenotypic data. Gene perturbation screens, omic datasets or high-throughput growth studies can

provide much more information, in particular with regards to layers of interactions that operate above or below metabolism itself. Generation of this data could improve the metabolic models reviewed in Chapter 3. Considering the minimal validation applied to the metabolic model of *M. capsulatus* in Chapter 5, the integration of more extensive experimental data also represents the next step in this model's development.

In addition to validation, another step that has been identified as crucial is the routine application of quality control when reconstructing a genome-scale metabolic model. The iterative, multistep process lends itself to human error, in addition to systemic errors introduced through automated approaches and generic biochemical databases. Again looking for computational precision and prowess to aid in this process, we presented memote in Chapter 4.

Automatic testing of each step in the reconstruction protocol provides clarity by identifying issues early. Version control contributes loss prevention and transparency by safekeeping each change and allowing users to distinguish contributions over the course of a project's history.

The introduction of memote, which both comprises a library of tests and the framework to execute them automatically, is not entirely devoid of associated issues either. On the one hand, the tests ought to be general enough to accommodate all possible representations of models, which have organically-grown drom of independent groups. On the other hand, the tests need to be precise enough to separate actual mistakes from tool-, or approach-specific idiosyncrasies. On top of that, the tool needs to cater to a wide audience with greatly varying backgrounds. A trade-off that achieves this will require a flexible and transparent approach both from the community and us, the developers.

Memote and the metabolic model for *M. capsulatus* both represent practical tools that benefit their respective areas. Memote introduces unbiased, centralized quality control and quality assurance to the COBRA community, while the metabolic model represents a tool for hypothesis-driven research in methanotrophy. Taken

together they can be employed in the rational development of one-carbon based cell factories for an improved production of single-cell protein. By extension, they hopefully contribute to the aversion of the grim prospect of climate-change induced food scarcity.

References For Outline, Chapters 1-3 & Conclusion

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., & Nielsen, J. (2013). The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. *PLoS Computational Biology*, *9*(3), e1002980. http://doi.org/10.1371/journal.pcbi.1002980

Akberdin, I. R., Thompson, M., Hamilton, R., Desai, N., Alexander, D., Henard, C. A., … Kalyuzhnaya, M. G. (2018). Methane utilization in Methylomicrobium alcaliphilum 20ZR: a systems approach. *Scientific Reports*, *8*(1), 2512. http://doi.org/10.1038/s41598-018-20574-z

Anthony, C. (1983). *The biochemistry of methylotrophs*. *London: Academic* (Vol. 75). http://doi.org/10.1016/0300-9629(83)90116-0

Anthony, C. (2011). How half a century of research was required to understand bacterial growth on C1 and C2 compounds; the story of the serine cycle and the ethylmalonyl-CoA pathway. *Science Progress*. http://doi.org/10.3184/003685011X13044430633960

Ates, Ö., Oner, E. T., & Arga, K. Y. (2011). Genome-scale reconstruction of metabolic network for a halophilic extremophile, Chromohalobacter salexigens DSM 3043. *BMC Systems Biology*, *5*(1), 12. http://doi.org/10.1186/1752-0509-5-12

Aung, H. W., Henry, S. A., & Walker, L. P. (2013). Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Industrial Biotechnology*, *9*(4), 215–228. http://doi.org/10.1089/ind.2013.0013

Balasubramanian, R., Smith, S. M., Rawat, S., Yatsunyk, L. A., Stemmler, T. L., & Rosenzweig, A. C. (2010). Oxidation of methane by a biological dicopper centre. *Nature*, *465*(7294), 115–119. http://doi.org/10.1038/nature08992

BASF. (2013). BASF produces first commercial volumes of butanediol from renewable raw material. Retrieved February 26, 2018, from https://www.basf.com/en/company/news-and-media/news-releases/2013/11/p-13-538.html

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., … Zhang, J. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. http://doi.org/10.1093/nar/gkw1099

Belin, B. J., Busset, N., Giraud, E., Molinaro, A., Silipo, A., & Newman, D. K. (2018). Hopanoid lipids: from membranes to plant–bacteria interactions. *Nature Reviews Microbiology*. http://doi.org/10.1038/nrmicro.2017.173

Benedict, M. N., Gonnerman, M. C., Metcalf, W. W., & Price, N. D. (2012). Genome-scale metabolic reconstruction and hypothesis testing in the

methanogenic archaeon Methanosarcina acetivorans C2A. *Journal of Bacteriology*, *194*(4), 855–865. http://doi.org/10.1128/JB.06040-11

Bennett, R. K., Steinberg, L. M., Chen, W., & Papoutsakis, E. T. (2018). Engineering the bioconversion of methane and methanol to fuels and chemicals in native and synthetic methylotrophs. *Current Opinion in Biotechnology*, *50*, 81–93. http://doi.org/10.1016/j.copbio.2017.11.010

Bollinger, J. M., & Broderick, J. B. (2009). Frontiers in enzymatic C-H-bond activation. *Current Opinion in Chemical Biology*, *13*(1), 51–57. http://doi.org/10.1016/j.cbpa.2009.03.018

Brauman, K., Cassidy, E. S., Gerber, J., Foley, J. A., Ramankutty, N., Brauman, K. A., … Siebert, S. (2011). Solutions for a Cultivated Planet, (October). http://doi.org/10.1038/nature10452

Brindescu, C., Codoban, M., Shmarkatiuk, S., & Dig, D. (2014). How Do Centralized and Distributed Version Control Systems Impact Software Changes ?

But, S. Y., Khmelenina, V. N., Reshetnikov, A. S., Mustakhimov, I. I., Kalyuzhnaya, M. G., & Trotsenko, Y. A. (2015). Sucrose metabolism in halotolerant methanotroph Methylomicrobium alcaliphilum 20Z. *Archives of Microbiology*, *197*(3), 471–480. http://doi.org/10.1007/s00203-015-1080-9

Cal, A. J., Sikkema, W. D., Ponce, M. I., Franqui-Villanueva, D., Riiff, T. J., Orts, W. J., … Lee, C. C. (2016). Methanotrophic production of polyhydroxybutyrate-co-hydroxyvalerate with high hydroxyvalerate content. *International Journal of Biological Macromolecules*, *87*, 302–307. http://doi.org/10.1016/j.ijbiomac.2016.02.056

Camp, H. J. M. O. Den, Islam, T., Stott, M. B., Harhangi, H. R., Hynes, A., Schouten, S., … Dunfield, P. F. (2009). Minireview Environmental , genomic and taxonomic perspectives on methanotrophic Verrucomicrobia, *1*, 293–306. http://doi.org/10.1111/j.1758-2229.2009.00022.x

Campodonico, M. A., Andrews, B. A., Asenjo, J. A., Palsson, B. O., & Feist, A. M. (2014). Generation of an atlas for commodity chemical production in Escherichia coli and a novel pathway prediction algorithm, GEM-Path. *Metabolic Engineering*, *25*, 140–158. http://doi.org/10.1016/j.ymben.2014.07.009

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., … Karp, P. D. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, *42*. http://doi.org/10.1093/nar/gkt1103

Chacon, S., & Straub, B. (2014). *Pro Git* (2nd ed.). Apress. Retrieved from https://git-scm.com/book/en/v2

Chang, H. N. (2018). Software Applications for Phenotype Analysis and Strain,

771–791.

Cho, A., Yun, H., Park, J. H., Lee, S. Y., & Park, S. (2010). Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Systems Biology*, *4*. http://doi.org/10.1186/1752-0509-4-35

Ciais, P., Sabine, C., & Bala, G. (2014). Carbon and other biogeochemical cycles. *Global Climate Change 2014*, 465–469. http://doi.org/10.1017/CBO9781107415324.015

Clemente, J. (2017). Why U.S. Natural Gas Prices Will Remain Low.

Clomburg, J. M., Crumbley, A. M., & Gonzalez, R. (2017). Industrial biomanufacturing: The future of chemical production. *Science*, *355*(38). http://doi.org/10.1126/science.aag0804

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., … De Hoon, M. J. L. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, *25*(11), 1422–1423. http://doi.org/10.1093/bioinformatics/btp163

Coleman, W. J., VIDANES, G. M., Cottarel, G., Muley, S., KAMIMURA, R., JAVAN, A. F., … Groban, E. S. (2017). Biological Production of Multi-Carbon Compounds from Methane. Google Patents. Retrieved from https://www.google.com/patents/US20170152529

Comer, A. D., Long, M. R., Reed, J. L., & Pfleger, B. F. (2017). Flux balance analysis indicates that methane is the lowest cost feedstock for microbial cell factories. *Metabolic Engineering Communications*, *5*(July), 26–33. http://doi.org/10.1016/j.meteno.2017.07.002

Conrad, R. (2009). The global methane cycle: Recent advances in understanding the microbial processes involved. *Environmental Microbiology Reports*, *1*(5), 285–292. http://doi.org/10.1111/j.1758-2229.2009.00038.x

Cotten, C., & Reed, J. L. (2016). *Applications of Constraint-Based Models for Biochemical Production. Biotechnology for Biofuel Production and Optimization*. Elsevier B.V. http://doi.org/10.1016/B978-0-444-63475-7.00008-X

Dalton, H. (2005). The Leeuwenhoek Lecture 2000 The natural and unnatural history of methane-oxidizing bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1458), 1207–1222. http://doi.org/10.1098/rstb.2005.1657

Dash, S., Ng, C. Y., & Maranas, C. D. (2016). Metabolic Modeling of Clostridia : Current Developments and Applications, (January), 1–22. http://doi.org/10.1093/femsle/fnw004

de la Torre, A., Metivier, A., Chu, F., Laurens, L. M. L., Beck, D. A. C., Pienkos, P. T., … Kalyuzhnaya, M. G. (2015). Genome-scale metabolic reconstructions and theoretical investigation of methane conversion in

Methylomicrobium buryatense strain 5G(B1). *Microbial Cell Factories*, *14*(1), 188. http://doi.org/10.1186/s12934-015-0377-3

Dedysh, S. N., Khmelenina, V. N., Suzina, N. E., Trotsenko, Y. A., Semrau, J. D., Liesack, W., & Tiedje, J. M. (2018). Methylocapsa acidiphila gen . nov ., sp . nov ., a novel methane-oxidizing and dinitrogen-fixing acidophilic bacterium from Sphagnum bog, (2002), 251–261.

del Cerro, C., García, J. M., Rojas, A., Tortajada, M., Ramón, D., Galán, B., … García, J. L. (2012). Genome sequence of the methanotrophic poly-β-hydroxybutyrate producer Methylocystis parvus OBBP. *Journal of Bacteriology*, *194*(20), 5709–5710. http://doi.org/10.1128/JB.01346-12

Dong, T., Fei, Q., Genelot, M., Smith, H., Laurens, L. M. L., Watson, M. J., & Pienkos, P. T. (2017). A novel integrated biorefinery process for diesel fuel blendstock production using lipids from the methanotroph, Methylomicrobium buryatense. *Energy Conversion and Management*, *140*, 62–70. http://doi.org/10.1016/j.enconman.2017.02.075

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, *7*(1), 1. http://doi.org/10.1186/1752-0509-7-74

Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., K., … J.C. Minx. (2014). IPCC, 2014: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*.

Elliott, S. J., Zhu, M., Tso, L., Nguyen, H. H. T., Yip, J. H. K., & Chan, S. I. (1997). Regio- and stereoselectivity of particulate methane monooxygenase from Methylococcus capsulatus (Bath). *Journal of the American Chemical Society*, *119*(42), 9949–9955. http://doi.org/10.1021/ja971049g

Erb, T. J., Jones, P. R., & Bar-Even, A. (2017). Synthetic metabolism: metabolic engineering meets enzyme design. *Current Opinion in Chemical Biology*, *37*, 56–62. http://doi.org/10.1016/j.cbpa.2016.12.023

Ettwig, K. F., Butler, M. K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M. M. M., … Strous, M. (2010). Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature*, *464*(7288), 543–548. http://doi.org/10.1038/nature08883

Fiehn, O., Barupal, D. K., & Kind, T. (2011). Extending biochemical databases by metabolomic surveys. *Journal of Biological Chemistry*, *286*(27), 23637–23643. http://doi.org/10.1074/jbc.R110.173617

Field, C. B., Barros, V. R., Dokken, D. J., Mach, K. J., Mastrandrea, M. D., Bilir, T. E., … White, L. L. (2014). *2014: Summary for policymakers. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects.*

Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*.

Flynn, J. D., Hirayama, H., Sakai, Y., Dunfield, P. F., Klotz, M. G., Knief, C., … Kalyuzhnaya, M. G. (2016). Draft Genome Sequences of Gammaproteobacterial Methanotrophs Isolated from Marine Ecosystems. *Ncbi*, *4*(1), 1–2. http://doi.org/10.1128/genomeA.01629-15.Copyright

Fowler, M. (2006). Continuous Integration. Retrieved February 23, 2018, from https://www.martinfowler.com/articles/continuousIntegration.html

Fritzemeier, C. J., Hartleb, D., Szappanos, B., Papp, B., & Lercher, M. J. (2017). Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLoS Computational Biology*, *13*(4), 1–14. http://doi.org/10.1371/journal.pcbi.1005494

Furutani, M., Uenishi, A., & Iwasa, K. (2015). Recombinant cell, and method for producing 1,4-butanediol. Google Patents. Retrieved from https://www.google.dk/patents/US20150368677

Ge, X., Yang, L., Sheets, J. P., Yu, Z., & Li, Y. (2014). Biological conversion of methane to liquid fuels: Status and opportunities. *Biotechnology Advances*. http://doi.org/10.1016/j.biotechadv.2014.09.004

GlobalCarbonProject. (2016). Global Methane Budget. Retrieved February 27, 2018, from http://www.globalcarbonproject.org/methanebudget/16/files/MethaneInfographic2016.png

Gonzalez, J. E., Bennett, R. K., Papoutsakis, E. T., & Antoniewicz, M. R. (2018). Methanol assimilation in Escherichia coli is improved by co-utilization of threonine and deletion of leucine-responsive regulatory protein. *Metabolic Engineering*, *45*(December 2017), 67–74. http://doi.org/10.1016/j.ymben.2017.11.015

Gou, Z., Xing, X. H., Luo, M., Jiang, H., Han, B., Wu, H., … Zhang, F. (2006). Functional expression of the particulate methane mono-oxygenase gene in recombinant Rhodococcus erythropolis. *FEMS Microbiology Letters*, *263*(2), 136–141. http://doi.org/10.1111/j.1574-6968.2006.00363.x

Goyal, N., Widiastuti, H., Karimi, I. A., & Zhou, Z. (2014). A genome-scale metabolic model of Methanococcus maripaludis S2 for CO $_2$ capture and conversion to methane. *Mol. BioSyst.*, *10*(5), 1043–1054. http://doi.org/10.1039/C3MB70421A

Goyal, N., Widiastuti, H., Karimi, I. A., & Zhou Zhi, G. (2013). *Genome-scale metabolic network reconstruction and in silico analysis of Methanococcus maripaludis S2*. *Computer Aided Chemical Engineering* (Vol. 32). Elsevier B.V. http://doi.org/10.1016/B978-0-444-63234-0.50031-2

Hamilton, J. J., & Reed, J. L. (2014). Software platforms to facilitate reconstructing genome-scale metabolic networks. *Environmental Microbiology*, *16*(1), 49–59. http://doi.org/10.1111/1462-2920.12312

Hatfield, S., Subramanian, A., Palmer, T., & Düben, P. (2017). Improving weather forecast skill through reduced precision data assimilation. *Monthly Weather Review*, MWR-D-17-0132.1. http://doi.org/10.1175/MWR-D-17-0132.1

Haynes, C. A., & Gonzalez, R. (2014). Rethinking biological activation of methane and conversion to liquid fuels. *Nature Chemical Biology*, *10*(5), 331–339. http://doi.org/10.1038/nchembio.1509

Heavner, B. D., & Price, N. D. (2015). Comparative Analysis of Yeast Metabolic Network Models Highlights Progress, Opportunities for Metabolic Reconstruction. *PLoS Computational Biology*, *11*(11), 1–26. http://doi.org/10.1371/journal.pcbi.1004530

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., … Fleming, R. M. T. (2017). Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. *arXiv:1710.04038*, (October). http://doi.org/10.1038/protex.2011.234

Henard, C. A., Smith, H., Dowe, N., Kalyuzhnaya, M. G., Pienkos, P. T., & Guarnieri, M. T. (2016). Bioconversion of methane to lactate by an obligate methanotrophic bacterium. *Scientific Reports*, *6*(February), 1–9. http://doi.org/10.1038/srep21585

Henard, C. A., Smith, H. K., & Guarnieri, M. T. (2017). Phosphoketolase overexpression increases biomass and lipid yield from methane in an obligate methanotrophic biocatalyst. *Metabolic Engineering*, *41*(December 2016), 152–158. http://doi.org/10.1016/j.ymben.2017.03.007

Henry, C. S., Dejongh, M., Best, A. A., Frybarger, P. M., Linsay, B., & Stevens, R. L. (2010). High-throughput generation , optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*, *28*(9), 969–974. http://doi.org/10.1038/nbt.1672

Hill, E. A., Chrisler, W. B., Beliaev, A. S., & Bernstein, H. C. (2017). A flexible microbial co-culture platform for simultaneous utilization of methane and carbon dioxide from gas feedstocks. *Bioresource Technology*, *228*, 250–256. http://doi.org/10.1016/j.biortech.2016.12.111

Hopcroft, P. (2017). Ancient ice and the global methane cycle. *Nature*, *548*(7668), 403–404. http://doi.org/10.1038/548403a

Jahng, D., & Wood, T. K. (1994). Trichloroethylene and chloroform degradation by a recombinant pseudomonad expressing soluble methane monooxygenase from Methylosinus trichosporium OB3b. *Applied and Environmental Microbiology*, *60*(7), 2473–2482.

Jensen, C. (2007). *Engineering Drawing & Design*. *Engineering Drawing & Design*.

Retrieved from http://ewhighered.mcgraw-hill.com/sites/dl/free/0073521515/558630/chapter29.pdf

Kalyuzhnaya, M. G., Yang, S., Rozova, O. N., Smalley, N. E., Clubb, J., Lamb, A., … Lidstrom, M. E. (2013). Highly efficient methane biocatalysis revealed in a methanotrophic bacterium. *Nature Communications*, *4*(May), 1–7. http://doi.org/10.1038/ncomms3785

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353–D361. http://doi.org/10.1093/nar/gkw1092

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014). Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Research*, *42*. http://doi.org/10.1093/nar/gkt1076

Karp, P. D., Paley, S. M., Krummenacker, M., Latendresse, M., Dale, J. M., Lee, T. J., … Caspi, R. (2010). Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, *11*(1), 40–79. http://doi.org/10.1093/bib/bbp043

Karthikeyan, O. P., Chidambarampadmavathy, K., Cirés, S., & Heimann, K. (2015). Review of Sustainable Methane Mitigation and Biopolymer Production. *Critical Reviews in Environmental Science and Technology*, *45*(15), 1579–1610. http://doi.org/10.1080/10643389.2014.966422

Keppler, F., Hamilton, J. T. G., Braß, M., & Röckmann, T. (2006). Methane emissions from terrestrial plants under aerobic conditions. *Nature*, *439*(7073), 187–191. http://doi.org/10.1038/nature04420

Kim, W. J., Kim, H. U., & Lee, S. Y. (2017). Current state and applications of microbial genome-scale metabolic models. *Current Opinion in Systems Biology*, *2*, 10–18. http://doi.org/10.1016/j.coisb.2017.03.001

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., … Zeng, G. (2013). Three decades of global methane sources and sinks. *Nature Geoscience*, *6*(10), 813–823. http://doi.org/10.1038/ngeo1955

Kits, K. D., Klotz, M. G., & Stein, L. Y. (2015). Methane oxidation coupled to nitrate reduction under hypoxia by the Gammaproteobacterium Methylomonas denitrificans, sp. nov. type strain FJG1. *Environmental Microbiology*, *17*(9), 3219–3232. http://doi.org/10.1111/1462-2920.12772

Lawton, T. J., & Rosenzweig, A. C. (2016a). Biocatalysts for methane conversion: big progress on breaking a small substrate. *Current Opinion in Chemical Biology*, *35*, 142–149. http://doi.org/10.1016/j.cbpa.2016.10.001

Lawton, T. J., & Rosenzweig, A. C. (2016b). Methane-Oxidizing Enzymes: An Upstream Problem in Biological Gas-to-Liquids Conversion. *Journal of the American Chemical Society*, *138*(30), 9327–9340. http://doi.org/10.1021/jacs.6b04568

Leak, D. J., & Dalton, H. (1986a). Growth yields of methanotrophs - 1. Effect of copper on the energetics of methane oxidation. *Applied Microbiology and Biotechnology*, *23*(6), 470–476. http://doi.org/10.1007/BF02346062

Leak, D. J., & Dalton, H. (1986b). Growth yields of methanotrophs 2. A theoretical analysis. *Applied Microbiology and Biotechnology*, *23*(6), 477–481. http://doi.org/10.1007/BF02346063

Lieven, C., Gernaey, K. V., Herrgard, M. J., & Sonnenschein, N. (2018). Modeling Methanotrophy: A genome-scale reconstruction of Methylococcus capsulatus. *Manuscript in Preparation*.

Lionetti, G. (2012). What is version control: centralized vs. DVCS. Retrieved February 23, 2018, from https://www.atlassian.com/blog/software-teams/version-control-centralized-dvcs

Liu, J., Chen, H., Zhu, Q., Shen, Y., Wang, X., Wang, M., & Peng, C. (2015). A novel pathway of direct methane production and emission by eukaryotes including plants, animals and fungi: An overview. *Atmospheric Environment*, *115*, 26–35. http://doi.org/10.1016/j.atmosenv.2015.05.019

Machado, D., Herrgård, M. J., & Rocha, I. (2016). Stoichiometric Representation of Gene–Protein–Reaction Associations Leverages Constraint-Based Analysis from Reaction to Gene-Level Phenotype Prediction. *PLoS Computational Biology*, *12*(10), 1–24. http://doi.org/10.1371/journal.pcbi.1005140

Manglaviti, M., Coronado-montoya, E., Gallaba, K., & Mcintosh, S. (2017). An Empirical Study of the Personnel Overhead of Continuous Integration, 14–17. http://doi.org/10.1109/MSR.2017.31

Mardina, P., Li, J., Patel, S. K. S., Kim, I. W., Lee, J. K., & Selvaraj, C. (2016). Potential of immobilized whole-cell Methylocella tundrae as a biocatalyst for methanol production from methane. *Journal of Microbiology and Biotechnology*, *26*(7), 1234–1241. http://doi.org/10.4014/jmb.1602.02074

Masui, T., Matsumoto, K., Hijioka, Y., Kinoshita, T., Nozawa, T., Ishiwatari, S., … Kainuma, M. (2011). An emission pathway for stabilization at 6 Wm-2radiative forcing. *Climatic Change*, *109*(1), 59–76. http://doi.org/10.1007/s10584-011-0150-5

McAnulty, M. J., Poosarla, V. G., Li, J., Soo, V. W. C., Zhu, F., & Wood, T. K. (2017). Metabolic engineering of Methanosarcina acetivorans for lactate production from methane. *Biotechnology and Bioengineering*, *114*(4), 852–861. http://doi.org/10.1002/bit.26208

McCloskey, D., Palsson, B., & Feist, A. M. (2013). Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Molecular Systems Biology*, *9*(1), 1–15. http://doi.org/10.1038/msb.2013.18

Meyer, M. (2014). Continuous Integration and Its Tools, 14–16.

Milne, C. B., Kim, P. J., Eddy, J. A., & Price, N. D. (2009). Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology. *Biotechnology Journal*, *4*(12), 1653–1670. http://doi.org/10.1002/biot.200900234

Mohammadi, S., Pol, A., Alen, T. A. Van, & Jetten, M. S. M. (2016). Methylacidiphilum fumariolicum SolV , a thermoacidophilic " Knallgas " methanotroph with both an oxygen-sensitive and -insensitive hydrogenase, *11*(4), 945–958. http://doi.org/10.1038/ismej.2016.171

Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., … Palsson, B. O. (2017). iML1515, a knowledgebase that computes Escherichia coli traits. *Nature Biotechnology*, *35*(10), 904–908. http://doi.org/10.1038/nbt.3956

Monk, J., Nogales, J., & Palsson, B. O. (2014). Optimizing genome-scale network reconstructions. *Nature Biotechnology*, *32*(5), 447–452. http://doi.org/10.1038/nbt.2870

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., … Zhang, H. (2013). Anthropogenic and Natural Radiative Forcing. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 659–740. http://doi.org/10.1017/ CBO9781107415324.018

Nagappan, N., & Maximilien, E. M. (2008). Realizing quality improvement through test driven development : results and experiences of four industrial teams, 289–302. http://doi.org/10.1007/s10664-008-9062-z

Nazem-Bokaee, H., Gopalakrishnan, S., Ferry, J. G., Wood, T. K., & Maranas, C. D. (2016). Assessing methanotrophy and carbon fixation for biofuel production by Methanosarcina acetivorans. *Microbial Cell Factories*, *15*(1), 1–13. http://doi.org/10.1186/s12934-015-0404-4

Ng, C. Y., Khodayari, A., Chowdhury, A., & Maranas, C. D. (2015). Advances in de novo strain design using integrated systems and synthetic biology tools. *Current Opinion in Chemical Biology*, *28*, 105–114. http://doi.org/10.1016/j.cbpa.2015.06.026

Nguyen, A. D., Hwang, I. Y., Chan, J. Y., & Lee, E. Y. (2016). Reconstruction of methanol and formate metabolic pathway in non-native host for biosynthesis of chemicals and biofuels. *Biotechnology and Bioprocess Engineering*, *21*(4), 477–482. http://doi.org/10.1007/s12257-016-0301-7

Niftrik, L. Van. (2014). Expanding the Verrucomicrobial Methanotrophic World : Description of Three Novel Species of Methylacidimicrobium gen . nov ., *80*(21), 6782–6791. http://doi.org/10.1128/AEM.01838-14

Nyman, R., Kapadia, S., Tuckett, D., Gregory, D., Ormerod, P., Smith, R., … Smith, R. (2018). News and narratives in financial systems: exploiting big data for systemic risk assessment Staff Working Paper, (704).

O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., & Palsson, B. Ø. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, *9*, 693. http://doi.org/10.1038/msb.2013.52

Ochsner, A. M., Sonntag, F., Buchhaupt, M., Schrader, J., & Vorholt, J. A. (2014). Methylobacterium extorquens: methylotrophy and biotechnological applications. *Applied Microbiology and Biotechnology*, *99*(2), 517–534. http://doi.org/10.1007/s00253-014-6240-3

Orth, J. D., Palsson, B. Ø., & Fleming, R. M. T. (2010). Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide. *EcoSal Plus*, *4*(1). http://doi.org/10.1128/ecosalplus.10.2.1

Øverland, M., Tauson, A.-H., Shearer, K., & Skrede, A. (2010). Evaluation of methane-utilising bacteria products as feed ingredients for monogastric animals. *Archives of Animal Nutrition*, *64*(3), 171–189. http://doi.org/10.1080/17450391003691534

Palson, B. (2015). *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press.

Palsson, B. O. (2015). *Systems biology: Constraint-based reconstruction and analysis*. *Systems Biology: Constraint-Based Reconstruction and Analysis*. http://doi.org/10.1017/CBO9781139854610

Patel, S. K. S., Mardina, P., Sang-Young, K., Jung-Kul, L., & In-Won, K. (2016). Biological methanol production by a Type II methanotroph. *Journal of Microbiology and Bio*, *26*(4), 717–724. http://doi.org/10.4014/jmb.1502.02022

Patel, S. K. S., Selvaraj, C., Mardina, P., Jeong, J. H., Kalia, V. C., Kang, Y. C., & Lee, J. K. (2016). Enhancement of methanol production from synthetic gas mixture by Methylosinus sporium through covalent immobilization. *Applied Energy*, *171*, 383–391. http://doi.org/10.1016/j.apenergy.2016.03.022

Petersen, L. A. H., Villadsen, J., Jørgensen, S. B., & Gernaey, K. V. (2017). Mixing and mass transfer in a pilot scale U-loop bioreactor. *Biotechnology and Bioengineering*, *114*(2), 344–354. http://doi.org/10.1002/bit.26084

Peyraud, R., Schneider, K., Kiefer, P., Massou, S., Vorholt, J. A., & Portais, J.-C. (2011). Genome-scale reconstruction and system level investigation of the metabolic network of Methylobacterium extorquens AM1. *BMC Systems Biology*, *5*(1), 189. http://doi.org/10.1186/1752-0509-5-189

Pieja, A. J., Morse, M. C., & Cal, A. J. (2017). Methane to bioproducts: the future of the bioeconomy? *Current Opinion in Chemical Biology*, *41*(1), 123–131. http://doi.org/10.1016/j.cbpa.2017.10.024

Placzek, S., Schomburg, I., Chang, A., Jeske, L., Ulbrich, M., Tillack, J., & Schomburg, D. (2017). BRENDA in 2017: New perspectives and new tools

in BRENDA. *Nucleic Acids Research*, *45*(D1), D380–D388. http://doi.org/10.1093/nar/gkw952

Pratscher, J., Vollmers, J., Wiegand, S., Dumont, M. G., & Kaster, A.-K. (2018). Unravelling the Identity, Metabolic Potential and Global Biogeography of the Atmospheric Methane-Oxidizing Upland Soil Cluster α. *Environmental Microbiology*, *0*. http://doi.org/10.1111/1462-2920.14036

Puri, A. W., Owen, S., Chu, F., Chavkin, T., Beck, D. A. C., Kalyuzhnaya, M. G., & Lidstrom, M. E. (2015). Genetic tools for the industrially promising methanotroph Methylomicrobium buryatense. *Applied and Environmental Microbiology*, *81*(5), 1775–1781. http://doi.org/10.1128/AEM.03795-14

Ravikrishnan, A., & Raman, K. (2015). Critical assessment of genome-scale metabolic networks: The need for a unified standard. *Briefings in Bioinformatics*, *16*(6), 1057–1068. http://doi.org/10.1093/bib/bbv003

Reeburgh, W. S. (2013). Global Methane Biogeochemistry. *Treatise on Geochemistry: Second Edition*, *5*, 71–94. http://doi.org/10.1016/B978-0-08-095975-7.00403-4

Reed, J. L., Vo, T. D., Schilling, C. H., & Palsson, B. O. (2003). An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biology*, *4*(9), R54. http://doi.org/10.1186/gb-2003-4-9-r54

Reshetnikov, A. S., Khmelenina, V. N., Mustakhimov, I. I., Kalyuzhnaya, M., Lidstrom, M., & Trotsenko, Y. A. (2011). Diversity and phylogeny of the ectoine biosynthesis genes in aerobic, moderately halophilic methylotrophic bacteria. *Extremophiles*, *15*(6), 653–663. http://doi.org/10.1007/s00792-011-0396-x

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., … Rafaj, P. (2011). RCP 8.5-A scenario of comparatively high greenhouse gas emissions. *Climatic Change*, *109*(1), 33–57. http://doi.org/10.1007/s10584-011-0149-y

Richards, M. A., Lie, T. J., Zhang, J., Ragsdale, S. W., Leigh, J. A., & Price, N. D. (2016). Exploring hydrogenotrophic methanogenesis: A genome scale metabolic reconstruction of Methanococcus maripaludis. *Journal of Bacteriology*, *198*(24), 3379–3390. http://doi.org/10.1128/JB.00571-16

Ritala, A., Häkkinen, S. T., Toivari, M., & Wiebe, M. G. (2017). Single Cell Protein—State-of-the-Art, Industrial Landscape and Patents 2001–2016. *Frontiers in Microbiology*, *8*(October). http://doi.org/10.3389/fmicb.2017.02009

Rostkowski, K. H., Pfluger, A. R., & Criddle, C. S. (2013). Stoichiometry and kinetics of the PHB-producing Type II methanotrophs Methylosinus trichosporium OB3b and Methylocystis parvus OBBP. *Bioresource Technology*, *132*, 71–77. http://doi.org/10.1016/j.biortech.2012.12.129

Satish Kumar, V., Ferry, J. G., & Maranas, C. D. (2011). Metabolic reconstruction

of the archaeon methanogen Methanosarcina Acetivorans. *BMC Systems Biology*, *5*. http://doi.org/10.1186/1752-0509-5-28

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., … Zhu, Q. (2016). The global methane budget 2000-2012. *Earth System Science Data*, *8*(2), 697–751. http://doi.org/10.5194/essd-8-697-2016

Scheller, S., Yu, H., Chadwick, G. L., McGlynn, S. E., & Orphan, V. J. (2016). Artificial electron acceptors decouple archaeal methane oxidation from sulfate reduction. *Science*, *351*(6274), 703–707. http://doi.org/10.1126/science.aad7154

Schuster, S., Pfeiffer, T., & Fell, D. A. (2008). Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology*, *252*(3), 497–504. http://doi.org/10.1016/j.jtbi.2007.12.008

Shahin, M., Babar, M. A. L. I., & Zhu, L. (2017). Continuous Integration , Delivery and Deployment : A Systematic Review on Approaches , Tools , Challenges and Practices, (Ci), 3909–3943.

Sharpe, P. L., DiCosimo, D., Bosak, M. D., Knoke, K., Tao, L., Cheng, Q., & Ye, R. W. (2007). Use of transposon promoter-probe vectors in the metabolic engineering of the obligate methanotroph Methylomonas sp. strain 16a for enhanced C 40 carotenoid synthesis. *Applied and Environmental Microbiology*, *73*(6), 1721–1728. http://doi.org/10.1128/AEM.01332-06

Shima, S., Krueger, M., Weinert, T., Demmer, U., Kahnt, J., Thauer, R. K., & Ermler, U. (2012). Structure of a methyl-coenzyme M reductase from Black Sea mats that oxidize methane anaerobically. *Nature*, *481*(7379), 98–101. http://doi.org/10.1038/nature10663

Siegel, J. B., Smith, A. L., Poust, S., Wargacki, A. J., Bar-Even, A., Louw, C., … Baker, D. (2015). Computational protein design enables a novel one-carbon assimilation pathway. *Proceedings of the National Academy of Sciences*, 201500545. http://doi.org/10.1073/pnas.1500545112

Silverman, J., & PURCELL, T. J. (2014). Propylene synthesis using engineered enzymes. Google Patents. Retrieved from https://www.google.com/patents/WO2014047209A1?cl=en

Simeonidis, E., & Price, N. D. (2015). Genome-scale modeling for metabolic engineering. *Journal of Industrial Microbiology and Biotechnology*, *42*(3), 327–338. http://doi.org/10.1007/s10295-014-1576-3

Sirajuddin, S., & Rosenzweig, A. C. (2015). Enzymatic Oxidation of Methane. *Biochemistry*, *54*(14), 2283–2294. http://doi.org/10.1021/acs.biochem.5b00198

Smith, K. A., Ball, T., Conen, F., Dobbie, K. E., Massheder, J., & Rey, A. (2003). Exchange of greenhouse gases between soil and atmosphere: Interactions of soil physical factors and biological processes. *European Journal of Soil Science*,

54(4), 779–791. http://doi.org/10.1046/j.1351-0754.2003.0567.x

Soo, V. W. C., McAnulty, M. J., Tripathi, A., Zhu, F., Zhang, L., Hatzakis, E., … Wood, T. K. (2016). Reversing methanogenesis to capture methane for liquid biofuel precursors. *Microbial Cell Factories*, *15*(1), 11. http://doi.org/10.1186/s12934-015-0397-z

Springmann, M., Mason-D'Croz, D., Robinson, S., Garnett, T., Godfray, H. C. J., Gollin, D., … Scarborough, P. (2018). Global and regional health effects of future food production under climate change: a modelling study. *The Lancet*, *387*(10031), 1937–1946. http://doi.org/10.1016/S0140-6736(15)01156-3

Stein, L. Y., Yoon, S., Semrau, J. D., DiSpirito, A. A., Crombie, A., Murrell, J. C., … Klotz, M. G. (2010). Genome sequence of the obligate methanotroph Methylosinus trichosporium strain OB3b. *Journal of Bacteriology*, *192*(24), 6497–6498. http://doi.org/10.1128/JB.01144-10

Strong, P. J., Xie, S., & Clarke, W. P. (2015). Methane as a resource: Can the methanotrophs add value? *Environmental Science and Technology*, *49*(7), 4001–4018. http://doi.org/10.1021/es504242n

Subbian, E. (2017). Production of succinic acid from organic waste or biogas or methane using recombinant methanotrophic bacterium. Google Patents. Retrieved from https://www.google.dk/patents/US20170121740

Tao, L., Schenzle, A., & Odom, J. (2005). Novel carotenoid oxidase involved in biosynthesis of 4, 4′-diapolycopene dialdehyde. *Applied and Environmental Microbiology*, *71*(6), 3294–3301. http://doi.org/10.1128/AEM.71.6.3294

Tao, L., Sedkova, N., Yao, H., Ye, R. W., Sharpe, P. L., & Cheng, Q. (2007). Expression of bacterial hemoglobin genes to improve astaxanthin production in a methanotrophic bacterium Methylomonas sp. *Applied Microbiology and Biotechnology*, *74*(3), 625–633. http://doi.org/10.1007/s00253-006-0708-8

Thiele, I., & Palsson, B. Ø. (2010, January). *A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature protocols*. Nature Publishing Group. http://doi.org/10.1038/nprot.2009.203

Thomson, A. M., Calvin, K. V., Smith, S. J., Kyle, G. P., Volke, A., Patel, P., … Edmonds, J. A. (2011). RCP4.5: A pathway for stabilization of radiative forcing by 2100. *Climatic Change*, *109*(1), 77–94. http://doi.org/10.1007/s10584-011-0151-4

Timmers, P. H. A., Welte, C. U., Koehorst, J. J., Plugge, C. M., Jetten, M. S. M., & Stams, A. J. M. (2017). Reverse Methanogenesis and Respiration in Methanotrophic Archaea. *Archaea*. http://doi.org/10.1155/2017/1654237

Trotsenko, Y. A., & Murrell, J. C. (2008). Metabolic Aspects of Aerobic Obligate Methanotrophy, *63*(7), 5–6. http://doi.org/10.1016/S0065-2164(07)00005-6

Ugalde, U. O., & Castrillo, J. I. (2002). Single Cell Proteins from Fungi and Yeasts,

2, 123–149.

United Nations, Department of Economic and Social Affairs, P. D. (2017). World Population Prospects: The 2017 Revision, Key Findings and Advance Tables. *Working Paper No. ESA/P/WP/248*.

van Vuuren, D. P., Stehfest, E., den Elzen, M. G. J., Kram, T., van Vliet, J., Deetman, S., … van Ruijven, B. (2011). RCP2.6: Exploring the possibility to keep global mean temperature increase below 2°C. *Climatic Change*, *109*(1), 95–116. http://doi.org/10.1007/s10584-011-0152-3

Walpole, S. C., Prieto-Merino, D., Edwards, P., Cleland, J., Stevens, G., & Roberts, I. (2012). The weight of nations: An estimation of adult human biomass. *BMC Public Health*, *12*(1), 1. http://doi.org/10.1186/1471-2458-12-439

Weyant, J., Azar, C., Kainuma, M., & Kejun, J. (2009). Report of 2.6 versus 2.9 Watts/m2 RCPP evaluation panel. *Integrated Assessment* …. Retrieved from http://ipcc.ch/meetings/session17/doc3d.pdf%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Report+of+2.6+Versus+2.9+Watts/m2+RCPP+Evaluation+Panel#0

Whitaker, W. B., Jones, J. A., Bennett, R. K., Gonzalez, J. E., Vernacchio, V. R., Collins, S. M., … Papoutsakis, E. T. (2017). Engineering the biological conversion of methanol to specialty chemicals in Escherichia coli. *Metabolic Engineering*, *39*(September 2016), 49–59. http://doi.org/10.1016/j.ymben.2016.10.015

Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., … Müller, W. (2012). SABIO-RK--database for biochemical reaction kinetics. *Nucleic Acids Research*, *40*(Database issue), D790-6. http://doi.org/10.1093/nar/gkr1046

Wuebbles, D. J., & Hayhoe, K. (2002). Atmospheric methane and global change. *Earth-Science Reviews*, *57*(3–4), 177–210. http://doi.org/10.1016/S0012-8252(01)00062-9

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., … Van Dien, S. (2011). Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nature Chemical Biology*, *7*(7), 445–452. http://doi.org/10.1038/nchembio.580

Zhang, W., Ge, X., Li, Y. F., Yu, Z., & Li, Y. (2016). Isolation of a methanotroph from a hydrogen sulfide-rich anaerobic digester for methanol production from biogas. *Process Biochemistry*, *51*(7), 838–844. http://doi.org/10.1016/j.procbio.2016.04.003

Zhang, Y., Cai, J., Shang, X., Wang, B., Liu, S., Chai, X., … Wen, T. (2017). A new genome-scale metabolic model of Corynebacterium glutamicum and its application. *Biotechnology for Biofuels*, *10*(1), 1–16. http://doi.org/10.1186/s13068-017-0856-3

Zúñiga, C., Morales, M., Le Borgne, S., & Revah, S. (2011). Production of poly-β-hydroxybutyrate (PHB) by Methylobacterium organophilum isolated from a methanotrophic consortium in a two-phase partition bioreactor. *Journal of Hazardous Materials*, *190*(1–3), 876–882. http://doi.org/10.1016/j.jhazmat.2011.04.011

The Novo Nordisk Foundation
Center for Biosustainability
Bulding 220
Kemitorvet
2800 Kgs. Lyngby